



**HAL**  
open science

## Qu'a changé le numérique ?

Roxanne Comotti, Wendingoudi Émile Ouedraogo, Claire Wolfarth, Camille  
Noûs

### ► To cite this version:

Roxanne Comotti, Wendingoudi Émile Ouedraogo, Claire Wolfarth, Camille Noûs. Qu'a changé le numérique?. Colloque international des Etudiant×e×s chercheur×se×s en DIactique des langues et Linguistique, CEDIL'18, May 2018, Grenoble, France. 2020. hal-02641560

**HAL Id: hal-02641560**

<https://hal.univ-grenoble-alpes.fr/hal-02641560v1>

Submitted on 28 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License



Colloque international des  
Etudiant·e·s chercheur·se·s en  
Didactique des langues et  
Linguistique

## Actes du colloque **CEDIL'18**

---

*Qu'a changé le numérique ? Perspectives linguistiques et didactiques*

---

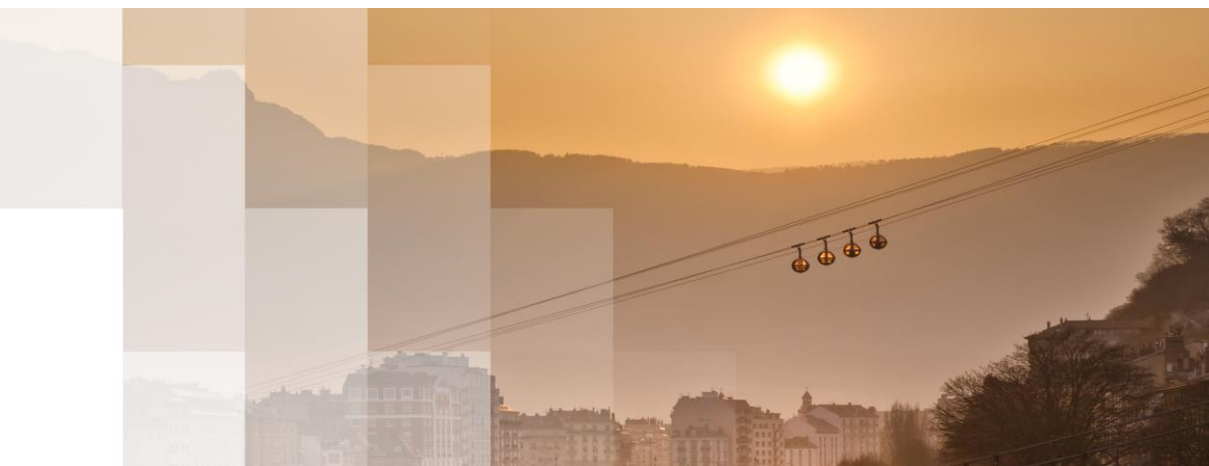
Coordonné par :

Roxanne COMOTTI

Wendingoudi Emile OUEDRAOGO

Claire WOLFARTH

Camille NOÛS



## TABLE DES MATIERES

Préambule .....	3
Comité d'organisation .....	5
<b>Le numérique dans la didactique des langues étrangères et maternelles .....</b>	<b>6</b>
Developing the UniON! mobile app to foster incidental learning of languages and cultures . 7 Antonella VALVA, Cristiana CERVINI	
Rapport à l'orthographe d'apprenants du secondaire et du supérieur dans les écrits numériques extrascolaires .....	17
Hélène LE LEVIER	
Understanding, setting up and evaluating a blended language course: a systemic approach .26 Aude LABETOULLE	
Corpus d'apprenants et recherche linguistique. Une étude sur l'italien. ....	37
Emanuele CASANI	
Approcher les textes par les notions de « genres » et de « types de discours » : choix didactiques pour la concrétisation au sein d'un environnement numérique .....	49
Triscia BIAGIOTTI, Giada QUARANTA	
<b>Le numérique dans les méthodologies de recherche.....</b>	<b>63</b>
Analyse de la combinatoire des verbes d'affect en arabe et en français fondée sur les corpus numériques <i>Emolex</i> et <i>Arabicorpus</i> .....	64
Safa ZOUAIDI	
Analyse bibliométrique des recherches sur l'écriture en langue seconde et étrangère en chine .....	73
Renlei WANG	
Le traitement de transcriptions d'écrits scolaires : un protocole d'annotations pour analyser les interventions des enseignants .....	85
Sara MAZZIOTTI	
Étude des formules expressives des interactions dans un corpus de tweets .....	94
Najwa GHARBI	
La concession en français dans des débats médiatiques .....	110
Thi Thu Trang DO	
<b>Le numérique dans les pratiques langagières .....</b>	<b>122</b>
Impact de la transition numérique sur la production de ressentis .....	123
Pauline SOUTRENON	
Néologie lexicale et médias sociaux : le cas des néologismes par abrègement de l'anglais.129 Pauline LEFEBVRE	

---

## PRÉAMBULE

Roxanne COMOTTI<sup>a</sup>, Wendingoudi Emile OUEDRAOGO<sup>a</sup>, Claire WOLFARTH<sup>a</sup>, Camille NOÛS<sup>b</sup>

[roxanne.comotti@univ-grenoble-alpes.fr](mailto:roxanne.comotti@univ-grenoble-alpes.fr), [wendingoudi-emile.ouedraogo@univ-grenoble-alpes.fr](mailto:wendingoudi-emile.ouedraogo@univ-grenoble-alpes.fr), [claire.wolfarth@univ-grenoble-alpes.fr](mailto:claire.wolfarth@univ-grenoble-alpes.fr)

<sup>a</sup>LIDILEM, Université Grenoble Alpes

38058 Grenoble cedex 9, France

<sup>b</sup>Laboratoire Cogitamus

France

---

Cet ouvrage fait suite au Colloque international des Étudiant·e·s chercheur·se·s en Didactique des langues et en Linguistique (CEDIL) qui a eu lieu du 30 mai au 1<sup>er</sup> juin 2018. Cette manifestation scientifique, qui se tient tous les quatre ans, est une opportunité d'échanges pour les jeunes chercheur·se·s de la communauté internationale. Elle leur permet d'entrer en contact avec les réseaux scientifiques de leurs domaines mais aussi de découvrir de nouvelles méthodologies de travail (modalités de recueil et de traitement de données, constitution et partage de corpus, etc.). CEDIL contribue également à la formation des jeunes chercheur·se·s tout en valorisant leurs contributions scientifiques. Cette quatrième édition a été consacrée à la thématique du numérique : *“Qu'a changé le numérique ? Perspectives linguistiques et didactiques”*. Cette thématique a été pensée de manière à inclure le plus largement possible des travaux touchant aux différentes disciplines des sciences du langage, ce qui a permis au colloque de s'ancrer au cœur des enjeux actuels du domaine, à savoir la constitution des outils numériques et les usages afférents. En effet, l'omniprésence de ces outils et ressources numériques dans le quotidien impacte le rapport à la langue. Elle en modifie l'usage et, par conséquent, interroge la recherche en sciences du langage tant dans les méthodes qu'elle met en œuvre que dans les objets qu'elle observe. Les communicants ont pu traiter de la place du numérique, de son application et de ses conséquences selon trois champs : les méthodologies de recherche, la didactique des langues étrangères et maternelles et les pratiques langagières.

L'édition 2018 du colloque CEDIL a donné lieu à 29 communications de doctorant·e·s et jeunes chercheur·se·s. À chaque début de demi-journée était associée une session “booster”, lors de laquelle chaque communicant·e était invité·e à résumer sa présentation en trois minutes afin d'éveiller la curiosité des auditeur·trices. Ces boosters ont permis à chacun et chacune d'avoir un aperçu du contenu du colloque, nonobstant l'organisation en sessions parallèles.

Chaque demi-journée était également introduite par une conférence plénière permettant d'approfondir un des axes du colloque. Lors de la première journée, Mathieu Avanzi (Sorbonne Université, STIH) a présenté le projet *“Cartographeur la variation régionale en français à l'heure du numérique”* qui consiste à documenter largement les variantes dialectales à partir d'une collecte de données participatives permettant d'affiner leur distribution géographique. La seconde journée, Françoise Boch (Université Grenoble Alpes, LIDILEM) a présenté *“La grammaire en couleur, passage au numérique”*, une démarche d'apprentissage inductive de la grammaire et de l'orthographe élaborée par M. Laurent, qui fait actuellement l'objet d'une transposition à distance (modules Grammortho). Puis, Georges-Louis Baron (Université Paris V René Descartes, EDA) s'est intéressé aux modes d'intégration et à leurs effets sur les compétences des apprenants dans sa conférence *“Les technologies en éducation au cours du temps. Quels effets sur les compétences des apprenants ? Une réflexion rétrospective”*. Enfin, Rachel Panckhurst (Université de Montpellier 3, PRAXILING) a interrogé le lien entre numérique et pratiques langagières à partir du contexte des SMS dans

une conférence intitulée “*Le discours numérique médié (DNM) et les pratiques scripturales contemporaines*”. Une table ronde rassemblant les conférencier·e·s autour de la question du numérique a permis de clôturer le colloque.

À la suite du colloque, les communicant·e·s ont été invités à produire un article détaillant leurs travaux. La présente publication fait état de ces articles, sélectionnés par un comité scientifique. Elle est composée de trois parties qui font écho aux trois axes du colloque.

Dans une première partie, les auteur·e·s questionnent l’introduction et/ou les conséquences du numérique en didactique des langues. **Valva & Cervini** traitent de l’apprentissage des langues via le développement d’une application. **Le Levier** s’intéresse au rapport à l’orthographe de collégiens, de lycéens et d’étudiants de BTS dans leurs écrits numériques. **Labetoulle**, quant à elle, analyse l’apport d’un dispositif entre cours présentiels et cours à distance pour l’apprentissage d’une langue étrangère. **Casani** s’appuie sur l’étude d’un corpus d’apprenants italophones pour décrire leurs compétences morphosyntaxiques. Enfin, **Biagiotti** propose l’utilisation d’une approche didactique par « genres de texte » et « types de discours » pour l’apprentissage des langues-cultures en s’appuyant sur un environnement numérique spécifique.

La deuxième partie s’intéresse aux différents usages du numérique dans les méthodologies de recherche en sciences du langage. **Zouaidi** s’appuie sur deux corpus francophones et arabophones pour analyser le lexique des affects à travers une étude sémantique et syntaxique. **Wang** propose une analyse bibliométrique à partir d’outils numériques des manuels d’apprentissage des langues secondaires et étrangères en Chine. **Mazzioti** s’intéresse à l’apport du numérique pour l’analyse des interventions des enseignant·e·s dans les copies d’élèves. **Gharbi** exploite un corpus de tweets pour proposer une description linguistique de formules expressives. Enfin, **Do** étudie les marqueurs de *concession* en français à partir de corpus de débats médiatiques.

La dernière partie est consacrée à l’impact du numérique sur différentes pratiques langagières. **Soutrenon** analyse, grâce au traitement automatique des langues, les écrits du ressenti à l’aune de la transition numérique. Enfin, **Lefebvre** étudie l’influence des facteurs du numérique dans les phénomènes d’abrègement.

Cette publication, nous l’espérons, propose un éclairage sur une problématique actuelle dans les recherches en sciences du langage et contribue à faire connaître les travaux de jeunes chercheur·se·s.

Bonne lecture à tou·te·s !

## COMITE D'ORGANISATION



Le comité d'organisation est composé majoritairement de jeunes chercheur·se·s du laboratoire *Lidilem* accompagnés par plusieurs maître·sse·s de conférence et professeur·se·s issu·e·s de ce même laboratoire.

- Abdramane Diarra (D)
- Anda Fournel (D)
- Cindy De Amaral (D)
- Claire Wolfarth (D)
- Claude Ponton (EC)
- Eman Abdulsadik (D)
- Irina Mukhlynina (D)
- Julie Sorba (EC)
- Laure Gardelle (EC)
- Luca Pallanti (D)
- Marie-Paule Jacques (EC)
- Marinette Matthey (EC)
- Olivier Kraif (EC)
- Qiao Li (D)
- Roxanne Comotti (D)
- Sandra Lagrange-Lanaspre (D)
- Virginie Zampa (EC)
- Wendingoudi Emile Ouedraogo (D)
- William N. Havard (D)

D : Doctorant·e·s et jeunes docteur·e·s

EC : Enseignant·e·s - Chercheur·se·s

**LE NUMERIQUE DANS LA DIDACTIQUE DES LANGUES  
ETRANGERES ET MATERNELLES**

---

## DEVELOPING THE UNION! MOBILE APP TO FOSTER INCIDENTAL LEARNING OF LANGUAGES AND CULTURES

Antonella VALVA<sup>a</sup>, Cristiana CERVINI<sup>b</sup>

[antonella.valva2@unibo.it](mailto:antonella.valva2@unibo.it), [cristiana.cervini@unibo.it](mailto:cristiana.cervini@unibo.it)

<sup>a</sup>*Department of Modern Languages, Literatures and Cultures, University of Bologna, Bologna, Italy*

<sup>b</sup>*Department of Interpretation and Translation, University of Bologna, Forlì, Italy*

---

### 1. About ILOCALAPP and UniOn!

This paper will describe the development of UniOn!, an application for mobile devices, produced within the ILOCALAPP project between October 2015 and August 2018, and launched thereafter in its finalised version. Even though the use of mobile technology for learning purposes is increasingly taken for granted in many contexts, it still represents a great challenge in terms of (re)conceiving education. As Traxler & Kukulka-Hulme (2016) point out, the context for mobile learning has changed and become more and more context-aware, due to the ubiquity of the social use of mobile devices. These topics were exploited during the ILOCALAPP project, and its main result – the UniOn! application for language and culture learning – represents a case study in this regard. We will begin this contribution with a description of the project and the application, before discussing in detail the design considerations. We will conclude by presenting the pilot phase during which the app was tested and revised before its final release. Throughout this paper, special attention will be paid to the collaborative approach, which characterised the design, the development and the testing of the app, and which led to the features of UniOn!'s current version.

#### 1.1. The ILOCALAPP project

ILOCALAPP – Incidentally Learning Other Cultures and Languages through an APP – was a 3-year (2015-2018) Erasmus+ project aimed at the production of UniOn!, a mobile application for supporting mobility students through the incidental learning of four cultures and languages: Finnish, Italian, Polish and Portuguese. The project was coordinated by the University of Bologna (Italy) and was carried out in collaboration with the Adam Mickiewicz University of Poznań (Poland), the University of Lapland of Rovaniemi, (Finland) and the Centre for Social Studies of the University of Coimbra (Portugal). The international and interdisciplinary composition of the project Consortium was a first key-aspect for the creation of the application, for which different kinds of expertise (language, cultural, inter-cultural, ICT, multimedia, digital design, etc.) were combined to develop this tool. UniOn! is based on the principles of incidental learning and offers to its users the possibility to get acquainted with local language and culture in a smart, creative way (Ceccherelli, Cervini, Magni, Mirri, Rocchetti, Salomoni & Valva, 2016). ILOCALAPP was aimed at supporting the international university students spending a period of time in the cities covered by the project (i.e. Bologna, Coimbra, Rovaniemi, Poznań). The goal was to integrate the local reality, both in terms of language acquisition and cultural awareness, by means of a smart mobile application which also conveyed practical information (Ceccherelli *et al.*, 2016). The app was rapidly enlarged to embrace other typologies of target users (e.g. high-school students, visiting scholars, tourists, migrants, among others), also opening up to new, differentiated uses (Valva, 2018).

UniOn! differs from the standard, common-known applications for language learning because learning is supposed to happen implicitly, as a by-product of other activities carried out by the users. Following the principles of incidental learning, the acquisition of language and culture relies on the learners' participation in the process of knowledge creation, in informal, formal



and non-formal settings (Cervini, Solovova, Jakkula & Ruta, 2016). The ILOCALAPP project offers app users the opportunity to get acquainted with local language and culture because they are able to interact with the host country and integrate into it. Thus, UniOn! fosters mobile and context-aware learning, taking advantage of the mobile devices in order to create learning experiences which exploit the richness and uniqueness of the learner's environment; it enables and encourages learners to capture aspects of the environment, considering the environment itself as a learning resource (Traxler & Kukulska-Hulme, 2016).

## 1.2. *The UniOn! application*

UniOn! is an application which works on both Android and iOS systems. It is equipped with geo-localisation and context-awareness functionalities; and integrates other mobile applications and commonly used social media (Mirri, Roccetti, Salomoni, Mambelli & Valva, 2017). The contents of UniOn! include both texts and multimedia (images, audio, videos), along with links to external resources, services, and web sites that are available throughout the application. UniOn! can be used almost entirely offline, as only links to external resources require a connection. Some app settings are customisable by the users (notably font size and notifications) who can also choose to include selected dictionaries, link to a dedicated Facebook group and activate a save-notes function (Valva, 2018).

The UniOn! application was created in four versions which are similar but not identical. There is one for each language/culture/city of the ILOCALAPP project. All include some common features, as well as some particularities, which were chosen by the single teams according to the specific needs and characteristics of each language/culture/city. In all four versions of UniOn! the contents are contextualised in the host city, so as to provide the users with language, culture and practical information (Valva, Mirri & Salomoni, 2018).

All contents of the app are organised in eight main categories, which are the same for the four languages/cultures/cities of the project (Valva, *et al.*, 2018). Each category is then composed of three-to-six thematic groups, or subcategories, and then again, each sub-category consists of several texts and insets. In other words, the contents of UniOn! stem out from a macro-level (the category), they branch out into intermediary sections (the sub-categories and the texts) and move to in-depth sections (the insets) (Valva, 2018). All topics are declined according to the socio-cultural features of the specific city: even though the main topics to be included were previously decided at consortium level and after relevant investigation among the stakeholders (see par.3 for details), the actual development of each version is very culture specific and it is inevitably influenced by the context. The eight main categories, which are common to the four versions, embrace the following fields of contents:

Main categories	Fields of contents
University Life	Studying, settling in, student life...
Getting around	Transportation, walking and cycling
Food & drink	Restaurants, cafés, canteens, bars
Worth seeing	Monuments, buildings, squares, parks
Entertainment	Cinema, theatre, museums, music
Lifestyle	Nature, sports, well-being
Services & needs	Health, housing, post, money
Shopping	Markets and supermarkets, shopping centres, design

**Table 1 – UniOn ! contents per category**

In addition to these eight categories, there is a ninth category called WordZ which includes the main vocabulary covered by the app and which daily shows the word (or the phrase) of the day.



**Image 1 – Screenshot of the menu page in the UniON! for Bologna**

The categories and subcategories menus allow for full accessibility of all the app contents without following any specific or suggested order. There are, however, other modalities for entering the app contents. First, they can be accessed through the notifications which are sent to users on the basis of their position. Secondly, a search function is included in the home page, making it possible to search for specific information.

In addition, for each category a game was included which can be accessed from a specific icon. The games created for UniOn! exploit three different typologies of format (matching, filling in the blanks, multiple choice). Even though the format is quite standard, the content of the games was explicitly developed with the aim of recalling the contents of the categories. The games can be played at any time, as an introduction to the category and/or on its way out, but this latter option is recommended in order to foster the memorisation of category's language contents (Valva, 2018).

The features described above result from intense and productive dialogue with the end-users, who proved to be crucial actors in the app development. Their contribution will be explained as part of the next section.

## **2. Theoretical and design considerations**

### **2.1. View of learning**

UniOn! is a smart mobile tool, which aims at facilitating learning by letting the users co-construct meaning and knowledge. It exploits the trend of mobile technologies, which are attracting new users and providing increasing capacity. In this way mobile technologies can contribute to the quality of education in so far as they can facilitate learning in formal and informal contexts, by creating an interactive learning environment with multiple contexts (Bachore, 2015). Moreover, learning through mobile technologies is likely to make boundaries between educational settings and life more permeable (MacDonald & Creanor, 2010). For example, following this direction, the contents included in UniOn! build upon the learner's ability to take advantage of and adapt to different kinds of input, both within and outside the application.

In addition, language and culture are given equal status in the app, and are even considered mutually dependent. This is in accordance with Kramsch (2003), according to whom language conveys information and provides knowledge with structure, culture is embedded in it. In this perspective, language both reflects and constructs cultural reality for speakers (Kramsch, 2003). Culture is communicated through language, and culture learning stems from interactive exchanges allowing for action and reflection (Levy, 2007).

Explicit language teaching (i.e. explanation of rules, structured exercises, correction and evaluation activities, etc) is not a part of the application. On the contrary, the language learning is meant to be incidental and to result from other activities (observing, interacting, problem solving).

In fact, the kind of learning fostered by the UniOn! app is experiential and based upon the mediation of learners with their local context. Such a mediation is aimed at the definition of personal learning environments, and also of personal learning goals, forms of practice, feedback, adaptation, collaboration and reflection (Laurillard, 2007). Moreover, the approach to learning conveyed by UniOn! is incidental, i.e. learning that is not planned nor intentional, but rather a by-product of other activities carried out by the app-users, and happening without learners being necessarily conscious of it (Kerka, 2000). In incidental learning, learners are in contact with factual information and they are engaged by that, they learn by doing and they interact with people, but also with tasks, knowledge and experience. In other words, in incidental learning learners interact with the context where learning is (incidentally) taking place. From this perspective, learning with UniOn! is context-aware insofar as the learning experience can exploit the situations in which the learner is engaged and the learner's environment becomes itself a learning resource. Contextualised learning is also personalised, and controlled by the learner, even though in a collaborative form, as learners can build shared knowledge (Traxler & Kukulska-Hulme, 2016).

All the above mentioned characteristics imply some challenges when planning and developing an educational app: mobile users are more likely to be distracted, mobile learning materials can be used in very different surrounding contexts; the small size of screen impacts upon the readability of texts; the information available should be easily retrievable in order to be easily applied in real-life situations, and so on. Therefore, it was particularly important to adopt design strategies that would foster implicit and incidental learning in the app (Cervini, 2018). In order to elucidate such strategies, the project teams decided to base the design and the development of the app on three core values:

- i) the app should be collaborative and participatory, that is to say that it calls for the active participation of the learner/user, who is encouraged to be in control of usage and learning;
- ii) the app should be integrated and situated, that is to say that it fosters authenticity and real-life situations and meaning that learning is context-aware and context-specific;
- iii) the app should focus on informal and incidental learning, that is to say that learning derives from other activities carried out by the learners/users (e.g. exploration of the city, observation of the location, interaction in the place) (Cervini *et al.*, 2016).

These three core values which accompanied the creation of UniOn! were sustained by– and, at the same time, incorporated in– the concept of “mobile”: mobile users, mobile devices, mobile learning are exploited and promoted by the app (Valva, 2018).

## **2.2. Role of the users**

In order to exploit incidental, mobile and situated learning, special attention was paid to the end-users, as engaging them from the very start of the design phases of an application impacts greatly on the success of an application once it is in use (Mirri, Roccetti & Salomoni, 2018). Therefore, the ILOCALAPP project teams adopted a collaborative design and followed a user-centred approach; the current features of the app are the result of a multi-stage end-users consultation.

The end-users consultation started at the very beginning of the project, including “traditional channels” (i.e. online surveys and face-to-face reviews) as well as some principles from user-

centred design approach and participatory methodology; which designing *for* the users was combined to designing *with* the users (Sanders, 2002). The end-users participated in experience-prototyping sessions and they were asked to test the different prototypes available at the different stages of production.

In the initial months of the project (January-May 2016), the end-users consultation consisted in: i) an online questionnaire, ii) focus groups and iii) experience prototyping. The online questionnaire was a large, international consultation conducted between February and March 2016; the questionnaire was sent to more than 25.000 mobility students and gathered 2.350 responses from all over Europe and the world. The survey collected information about the cultural areas of interest for mobility students, their views on using language & culture apps, and background information about the general usage of apps and mobile devices (Valva *et al.*, 2018). The results of the online questionnaire were exploited to better identify the nine categories of contents to be included in the app and the features which could help motivate and engage the users of language and culture apps (Mirri *et al.*, 2017). For the second consultation activity, the focus groups were specific, dedicated meetings with small groups of students (10-15 people), which were held in the Consortium institutions in April 2016. The focus groups aimed at getting feedback and comments from participants on the functionalities, the interactions, and the flow of the application, in terms of what, how and when offering contents and activities (Mirri *et al.*, 2017). Finally, the third preliminary consultation activity consisted in experience prototyping sessions, which were carried out in the Consortium institutions in May 2016. During the sessions, the end-users simulated the interaction with the app by carrying out some specific tasks (e.g. buying a ticket, finding a place, ordering in a café, etc.) using a prototype. After the sessions, they were asked about specific aspects of the interaction with the app prototype and their answers helped the project teams progress with the definition and design of the UniOn! app functionalities and services (Valva *et al.*, 2018).

After this initial consultation, more tests involving the end-users were conducted while the app prototypes were being updated. In January 2017, an initial assessment was carried out with a prototype including the first two categories of contents. The test was conducted with a small group of mobility students who raised relevant issues, such as the typology of contents to be included, the necessary functionalities of the app, the presence (or not) of a vehicular language. In September 2017, another test was conducted with a prototype including six out of the nine planned categories and almost all the planned functionalities. Again, relevant comments were collected, especially for the geo-localisation and the search functions.

Finally, the pilot phase of the UniOn! app took place in the initial months of 2018, when the app prototype was finalised (see below for details). It is worth underlining that all the consultation activities involving the end-users and their active participation in all the project stages drove the development of the application and had a positive impact on the final results.

### **2.3. Creation of app contents**

Before creating contents for incidental language learning, the authors focused upon the most important goals to achieve, as well as upon the constraints to be respected. Starting from the learning goals, and in line with experiential and situational incidental learning methodologies, we aimed at transferring the linguistic and cultural richness of the territory, through the descriptions of its places (institutional, cultural and/or fun), within the frame of a communicative action-oriented approach.

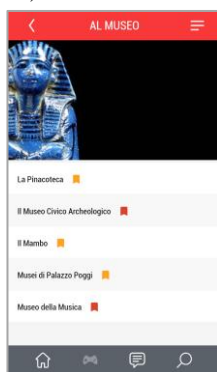
The app's contents are addressed to the whole community of mobility students, independently from their competence in the target language (the app is addressed to beginners as well as to more advanced students). This represented both a goal and a constraint, because it had an impact on the writing strategies adopted by the authors. A key constraint was small screen

reading features, in the case of reading texts on smartphones or other digital mobile devices. These considerations are further discussed below in relation to three key-questions:

1) *Given the wide variety of possibilities, which places would it be most appropriate to include in the contents?* Places to be geo-localised in the app had to be limited, so we had to choose carefully respecting the notion of authenticity, and the desire to promote the most interesting and unique places of the territory. Concerning authenticity, we tried to retrace the habits of international students, in their daily-life as university students, but also as young adults living a new life abroad. That is why in UniOn! both institutional places (university, libraries, public offices, etc.) and entertainment ones (cinemas, museums, pubs, etc.) are included.

2) *How can we propose texts and contents suitable for all types of public?* This was one of the most challenging issues in writing texts and in conceiving multimedia contents for the app. Learning is more likely to take place if linguistic-cultural inputs are rich and motivating, and if contents' complexity is coherent with students' language proficiency; complex texts can be demotivating for beginners and simple texts can be boring for advanced students. For these reasons, the authors developed a series of guiding principles:

i) The use of coloured icons to distinguish easy texts from those of medium or high difficulty. This suggests a reading path that best suits students' needs (green for easy texts, yellow for medium ones and red for difficult ones) and avoids disorientation;



**Image 2 – coloured icons to distinguish easy, intermediate and advanced texts.**

ii) the application of techniques for writing highly comprehensible texts. For example, in the Italian UniOn! for Bologna, short and coordinated sentences were preferred to long and subordinate clauses in texts identified as easy or medium. Moreover, the lexicon was carefully chosen, composed mainly of high-frequency words. Less frequent words, such as idiomatic expressions are usually followed by reformulations, glosses, and/or synonyms, to enhance understanding. Key-words and expressions that are worth remembering are hyperlinked to the related lexical sheet, and vice-versa, where words are organised in semantic area. This kind of consultation creates redundancy, without being too repetitive and boring because the same or similar inputs are proposed again in different contexts;

iii) the exploitation of the theories of multimodality, to enhance the multimodal exploration of contents: “as a phenomenon of communication, multimodality defines the combination of different semiotic resources, or modes, in texts and communicative events, such as still and moving image, speech, writing, layout, gesture, and/or proxemics” (Adami, 2015). A multimodal approach is coherent with context-aware and experiential learning, because students experience the authentic place while surfing the related contents in the app;

iv) the promotion of peer-tutoring practices among students, through the use of online collaborative tools: groups of students can experience similar difficulties or have similar

doubts; in this case background and competence diversity is a real advantage because, in more than one dedicated online space, students belonging to UniOn! community can share doubts and experiences.

In the event that all these comprehension strategies are not enough, texts are also proposed in a version translated into English.

3) *How can we promote integrated skills work, in an incidental way, while getting familiar with the places and with the typical habits of the places?* UniOn! intends to prod the development of integrated skills -receptive and interactive- in an implicit way. Written texts and lexical sheets accompany audio-video sources, presenting the same topic from different perspectives and favouring different cognitive styles.

In order to enhance students' active use of new lemmas and expressions and to blend authentic linguistic input read or heard in the places, we designed a specific feature in the app called *Tips to Talk*. This section is dedicated to the listening and reading of the most prototypical interactions in a place (i.e. questions & answers) and it can be filled in with new proposals intercepted during live oral exchanges.

Moreover, the real heteroglossia of the territories is the expression of the linguistic-cultural identities of citizens, shop owners, and businesses in general. The linguistic landscape can be exploited profitably for meaningful and incidental language learning while strolling around the places using UniOn!. "The language of public road signs, advertising billboards, street names, place names, commercial shop signs, and public signs on government buildings combines to form the linguistic landscape of a given territory, region, or urban agglomeration" (Cenoz, Groter 2006).

### 3. Piloting the application

The pilot phase of the UniOn! app took place in the four project cities in the initial months of 2018. In the project application, the pilot phase was defined as "the testing of the app by a group of selected students who will be provided with the tool and asked to use it for a defined period of their mobility experience" (ILOCALAPP Project website). The end- users' consultation of the first project phase and all the testing activities carried out at all project stages influenced the structure of the pilot phase and resulted in a tailor-made piloting of the final app prototype.

#### 3.1. Organisation of the pilot phase

In November 2017, the selection process of the pilot users started in the four institutions of the project Consortium. In the end, more than 100 students were included in the testing of the final prototype of the four versions of the app (about 25 per city). The majority of pilot users were university mobility students, as they represented the main target identified by the project teams. This group consisted of foreign students spending a period (a semester, a full year) in the university of one of the project cities. Other typologies of students were also included, for instance university students coming from different cities of the same country (who very much appreciated the cultural and practical information included in UniOn! even though they were not as interested in language information) or high-school students in their final years (who were similar to the major group of pilots in terms of age and cultural background). Pilot users with different language levels (from A1 to C2 and even native speakers) were accepted and welcomed, to assess the occurrence of incidental learning in different circumstances. Moreover, also pilot users using different devices and different operating systems (the app being compliant with both Android and iOS) were involved in order to test the app response within different platforms (Valva *et al.*, 2018).

From December 2017 to February 2018, the pilot users who described themselves as beginners in the target language were given the opportunity to take an online course to receive

preliminary information. The course remained at their disposal for the whole duration of the pilot phase, as a supporting tool. Only a few of them, however, took advantage of this opportunity. Moreover, they did not consider the two activities (taking an online language course and testing a mobile application about that same language and related culture) as complementary, as was intended by the Consortium teams.

In January 2018 nevertheless, an introductory seminar was organised in the Consortium institutions in order to inform all pilot users about the purposes of the piloting and provide them with all necessary information and tools to carry out the assessment. At this point, the pilot users were provided with an evaluation grid, which they were asked to use as a guideline to assess both the contents (in terms of interest, usefulness, readability and language learning efficacy) and the functionalities of the app.

After that seminar, the pilot phase started. The pilot users were free to use the app on their own to test it; and they were invited to move around the city with the app and to perform some tasks. A couple of meetings with pilot users were scheduled in each project institution during the pilot phase, to check progress and discuss eventual problems. After the introductory seminar, the participants had been asked to fill in a questionnaire about their expectations and background information, and in the end a final questionnaire was distributed to the pilot users, to collect their feedback. This one included questions about all the issues contained in the evaluation grid. In total, 92 pilot users amongst the Consortium returned the survey about their expectations and 80 students returned the final feedback form, showing that they had completed the testing activities (Valva *et al.*, 2018).

### **3.2. *Pilot users': Traits and feedback***

The pilot activities lasted until the end of February 2018, and thereafter the feedback was analysed. The preliminary questionnaire submitted by the participants in the introductory seminar provided background information about the pilot users themselves and about their expectations, including information about age, gender and nationality, but also about language background and experience with technology (Valva *et al.*, 2018). The data collected through this questionnaire provided a profile of the pilot users, the majority of them being females aged 19-22 years. The language background and the nationalities were quite diverse, with English being spoken by almost everybody at least as a second language. Fifty-six percent of respondents had already used an app for language learning. Their expectations concerning UniOn! were linked to the novelty of testing an innovative tool, and also to the possibility to learn new information and improve some skills. The discovery of their new context thanks to the app was also frequently mentioned in the submitted questionnaires as an expectation of the pilot users.

The evaluation grids which were distributed among the pilot users during their testing period focused on the contents and the functionalities of UniOn! They were meant to be relevant both for the pilot users, in so far as they were made aware of the issues to be checked, and for the project teams, as they served as guidelines to process the feedback received.

In the final evaluation questionnaire, after providing preliminary background information again, pilot users were invited to assess the items included in the evaluation grids. The evaluation questionnaire was structured in 16 Likert-scale questions, with a five-point agreement scale. The assessed items included topics, language, design, layout, geo-localisation, among others (Valva *et al.*, 2018).

The ILOCALAPP partners analysed the pilots' feedback in March 2018. The analysis of the results was done slightly differently within each team, as the feedback was specific to each of the four versions of the app. Moreover, the different typologies of users involved in the app testing (Erasmus students, students enrolled in international courses, national students, local

high school students, etc.) required the feedback to be carefully analysed, in order to contextualise it and verify its applicability.

The feedback confirmed that the topics in the app were considered useful and interesting and the texts understandable even though quite long. The games were regarded as useful to memorise information and in general the learning was perceived as quite intuitive. The layout was also considered intuitive, as the icons help identify the contents, and the design quite attractive. There were different opinions about the notifications related to the geo-localised contents: some pilot users felt that the system worked perfectly, while others reported some problems. Overall, the app was considered user-friendly.

The analysis of the pilots' feedback led to a plan for improving the app. In the Italian version, for instance, some texts were simplified slightly, some texts were added, and the indication of level was added to all texts of the Italian version through a system of coloured flags. The lexical notes were also improved and better organised, incorporating visual examples. The Tips to Talk were refined and new words to include in cat. 9 (WordZ) were selected. Moreover, final questions directly addressing the users were provided for almost all texts, to stimulate their reflection on the information provided. At the end of this revision phase, the finalised version UniOn! was released for all 4 languages and cultures involved in the project.

## 4. Conclusion

In this paper we have described the development of the UniOn! application within the ILOCALAPP project. The results collected during the project show that such an application interests both users and researchers. Further investigation is, however, needed in order to better understand the impact of the application upon the users, and notably the effects of mobile incidental learning through an app on measurable progress in language and/or culture learning. This is the focus of current research into real usage of the app, and it will be the object of future developments within the project teams.

### *Références bibliographiques*

- ADAMI, Elisabetta, (2016). Multimodality, in O. García, N. Flores and M. Spotti, *Oxford handbook of language and society*. Oxford: Oxford University Press.
- BACHORE, Mebratu Mulatu, (2015). Language Learning through Mobile Technologies: An Opportunity for Language Learners and Teachers, *Journal of Education and Practice*, vol.6, no. 31, 50-53.
- CECCHERELLI, Andrea, CERVINI, Cristiana, MAGNI, Elisabetta, MIRRI, Silvia, ROCCETTI, Marco, SALOMONI, Paola & VALVA, Antonella (2016). The ILOCALAPP project: a smart approach to language and culture acquisition, *The Future of Education Conference Proceedings*, 270-275.
- CENOZ, J., GROTER, D., (2006). Linguistic Landscape and Minority Languages. In *International Journal of Multilingualism*, Vol. 3, N° 1: 67-80.
- CERVINI, Cristiana (2018). Esperienze linguistico-culturali a Bologna: apprendere l'italiano L2 con UniOn!, in *Esperienze di e-learning per l'italiano*, VIALE, Matteo (Ed). Bologna: Bononia University Press - BUP.
- CERVINI, Cristiana, SOLOVOVA, Olga, JAKKULA, Annukka, & RUTA, Karolina. (2016). Mobile assisted language learning of less commonly taught languages: learning in an incidental and situated way through an app, *CALL communities and culture – Short papers from EUROCALL 2016*, 81-86.
- KERKA, Sandra (2000). Incidental Learning. Trends and Issues Alert No. 18, ERIC Publications.



- KRAMSCH, Claire (2003). From theory to practice and back again. In BYRAM, Michael & GRUNDY, Peter (Eds.) *Context and Culture in Language Teaching and Learning*. Clevedon: Multilingual Matters, 4-17.
- LAURILLARD, Diana (2007). Pedagogical forms for mobile learning. In Pachler, Norbert (ed), *Mobile learning: towards a research agenda*. London: WLE Centre, IoE, 153-175.
- LEVY, Mike (2007). Culture, culture learning and new technologies: towards a pedagogical framework, *Language learning & technology, volume 11, issue 2*, 104-127.
- MACDONALD, Janet & CREANOR, Linda (2010). *Learning With Online and Mobile Technologies: a Student Survival Guide*. Farnham, Surrey, England: Gower.
- MIRRI, Silvia, ROCCETTI, Marco & SALOMONI, Paola (2018). Collaborative design of software applications: the role of users, *Human-Centric Computing And Information Sciences, 2018, 8*, 1 – 20.
- MIRRI, Silvia, ROCCETTI, Marco, SALOMONI, Paola, MAMBELLI, Giacomo & VALVA, Antonella (2017). On the design of an app for foreign languages incidental learning, *IEEE Symposium on Computers and Communications (ISCC) Conference Proceedings*: 111-116.
- SANDERS, Elizabeth B.N. (2002). From user-centered to participatory design approaches. In FRASCARA, Jorge (Ed.), *Design and the Social Sciences*. London and New York: Taylor & Francis Books Limited, 1-8.
- TRAXLER, John & KUKULSKA-HULME, Agnes (2016). Introduction to the Next Generation of Mobile Learning, in *Mobile Learning: The Next Generation*, TRAXLER, John & KUKULSKA-HULME, Agnes (Eds). New York and London: Routledge, 1-10.
- VALVA, Antonella, MIRRI, Silvia, & SALOMONI, Paola (2018 in press). User centred design applied to an app for incidental learning of languages and cultures, *12th International Multi-Conference on Society, Cybernetics and Informatics (IMSCI 2018) Conference Proceedings*.
- VALVA, Antonella, (2018). User-centred design, participatory methodology and incidental learning of language-culture: how to test the effectiveness of this approach? The case study of the UniOn! app, *Goodtechs '18 Conference Proceedings*.

### ***Références aux sites Internet***

ILOCALAPP Project website, <http://www.ilocalapp.eu/>. Consulted in December 2018.

# RAPPORT A L'ORTHOGRAPHE D'APPRENANTS DU SECONDAIRE ET DU SUPERIEUR DANS LES ECRITS NUMERIQUES EXTRASCOLAIRES

Hélène LE LEVIER<sup>a</sup>

[helene-fleur.le-levier@ac-lille.fr](mailto:helene-fleur.le-levier@ac-lille.fr)

<sup>a</sup>LIDILEM Université Grenoble Alpes, Grenoble, France

---

Les pratiques d'écriture numériques, envois de textos et participation aux réseaux sociaux en particulier, ont modifié le rapport des plus jeunes générations à l'écriture en créant des espaces de dialogue écrits où se développent de nouveaux modes d'expression. L'opinion publique rend facilement responsable ces nouvelles formes d'écriture du recul, par ailleurs avéré (Manesse, Cogis, & Dorgans-Robineau, 2007 ; Andreu & Steinmetz, 2016), du niveau orthographique des jeunes francophones. La recherche a montré que ce lien de cause à effet semblait peu fondé même si, par ailleurs, l'originalité de ces nouvelles modalités d'écriture n'est pas remise en cause (Bernicot Josie, 2013 ; Joannidès, 2014).

Il paraît alors intéressant de se demander si ces jeunes scripteurs accordent une place spécifique à la norme orthographique dans leurs écrits numériques ou si elle n'est que le reflet de leur rapport global à la norme.

Nous apporterons d'abord quelques éléments théoriques concernant le rapport entre norme orthographique et écrits numériques puis nous présenterons le double recueil de données que nous avons exploité avant d'exposer les principaux résultats.

## 1. Norme orthographique et écrits numériques

### 1.1. *Des modalités d'écriture qui évoluent avec les nouvelles technologies*

Fayol et Jaffré (2014) soulignent que les évolutions récentes de la société, en particulier les développements d'internet et des technologies numériques, renforcent l'importance de l'écriture, intimement liée à la lecture, dans le quotidien d'une grande partie de la population. Des populations qui écrivaient autrefois peu sont aujourd'hui amenées à produire des écrits abondants suivant des modalités qui se rapprochent souvent plus de la conversation informelle que de l'écrit formel. Bernicot (2013) souligne d'ailleurs que l'on voit apparaître plusieurs registres de langue écrits comme on avait l'habitude de traiter plusieurs registres de langue oraux.

### 1.2. *La variation des écritures en fonction des supports*

Les outils numériques offrent des espaces de socialisation diversifiés pouvant aboutir à une forme de variation diaphasique (Gadet, 2003), un même scripteur modifiant sa façon d'écrire en fonction du contexte dans lequel il écrit. Le fait que les élèves les plus capables de respecter la norme orthographique en contexte scolaire puissent aussi être ceux qui pratiquent le plus de textismes (Bernicot, 2013) montre que les jeunes générations actuelles pratiquent déjà cette adaptation de l'écrit au contexte.

Joannidès souligne également dans sa thèse (2014, p 231 et suivantes) que les élèves de collègue qu'elle a interrogés insistent sur le fait que leur respect de l'orthographe dépend beaucoup du contexte dans lequel ils écrivent et en particulier de la personne à laquelle ils écrivent.

### 1.3. *Le rapport à la norme orthographique en français*

On sait que le rapport à la norme, en particulier la norme orthographique, est très fort en français (Paveau & Rosier, 2008) et que le poids social de cette norme se manifeste très tôt.

Millet, Lucci et Billiez ont montré (1990) que l'importance de respecter la norme orthographique apparaissait dans les discours des élèves dès l'école primaire et se retrouvait au collège et au lycée malgré la mise en place d'une distance critique chez certains scripteurs. Cependant, comme on l'a dit plus haut, le rapport à l'écrit s'est profondément modifié depuis cette étude du fait du développement des outils numériques. On peut donc se demander quel est le rapport à la norme orthographique de jeunes d'aujourd'hui, notamment lorsqu'on les interroge sur leurs écrits numériques.

## 2. Un double recueil de données

Les résultats que nous présentons ici s'appuient sur deux recueils de données réalisés dans l'académie de Grenoble en 2016 et dans l'académie de Lille en 2017.

### 2.1. *Entretiens avec des collégiens*

Dans le cadre de l'enquête Orthocol<sup>1</sup>, nous avons réalisé 131 entretiens de dix à vingt minutes auprès d'élèves de troisième, c'est-à-dire de la quatrième année du secondaire français. Ces entretiens se déroulaient en deux temps. Dans un premier temps, les élèves étaient interrogés sur les raisonnements qui les amenaient à produire telle ou telle forme orthographique à partir d'un texte de dictée préalablement écrit. Dans un second temps, nous les interrogeons sur leur rapport à l'orthographe, les valeurs qu'ils lui associaient et leurs pratiques d'écriture (Le Levier Hélène in Mortamet Clara dir, 2019). Nous nous intéresserons ici à ce qu'ils déclarent concernant leurs pratiques d'écriture numériques, notamment le respect ou non de l'orthographe dans les textos.

### 2.2. *Questionnaires et entretiens avec des élèves de STS*

La seconde enquête que nous exploiterons ici a été réalisée auprès d'élèves de trois sections de technicien (STS) supérieur différentes, c'est-à-dire auprès d'élèves préparant trois brevets de technicien supérieur différents (BTS) : la section assistant de manager, la section services informatiques aux organisations et la section tourisme. Ces trois sections ont été choisies parce qu'il s'agit de sections tertiaires où la maîtrise de l'écrit en français est considérée comme une compétence professionnelle même si l'importance de cette compétence est plus fondamentale dans les sections assistant de manager et tourisme que dans la section services informatiques aux organisations.

L'enquête réalisée en STS s'est inspirée de l'enquête menée auprès des élèves de troisième<sup>2</sup>. Ainsi les 65 entretiens transcrits que nous exploiterons ici suivaient le déroulement des entretiens décrits pour les élèves de troisième. Cependant, afin de compléter certaines données, nous avons associé à la dictée préalable un questionnaire abordant le rapport à l'orthographe des étudiants. Nous exploiterons ici deux questions auxquelles ont répondu les 178 élèves de STS :

- 1) Hors du lycée, vous arrive-t-il d'écrire :
  - Des lettres
  - Des cartes postales
  - Des listes (de courses, de choses à faire...)
  - Des petits mots

<sup>1</sup>Ce projet a été financé dans le cadre de l'appel « Alpes Grenoble Innovation Recherche » (AGIR) de l'Université Grenoble Alpes. Nous remercions les chefs d'établissement pour leur collaboration et les élèves pour leur participation.

<sup>2</sup>La dictée utilisée dans la première partie de l'enquête a été modifiée pour mieux s'adapter à l'âge des enquêtés. Après la dictée, les élèves de STS ont également rempli un questionnaire dont nous parlons juste après. Les entretiens ont suivi le même déroulement mais les entretiens de STS s'appuyaient sur les questionnaires dont nous ne disposons pas en troisième.

- Des sms
- Des mails
- Des messages sur des forums
- Des messages sur les réseaux sociaux

Autres :

2) Lorsque vous écrivez hors des contextes professionnel et scolaire, faites-vous attention à votre orthographe (orthographe des mots mais aussi accords, conjugaisons...) ?

- Oui toujours
- Oui parfois
- Rarement
- Jamais

Ces deux questions permettent d'aborder les pratiques d'écriture et l'importance accordée au respect de la norme orthographique chez les 178 étudiants ayant participé à l'enquête. Les entretiens permettent de préciser et développer ces points chez 65 de ces étudiants.

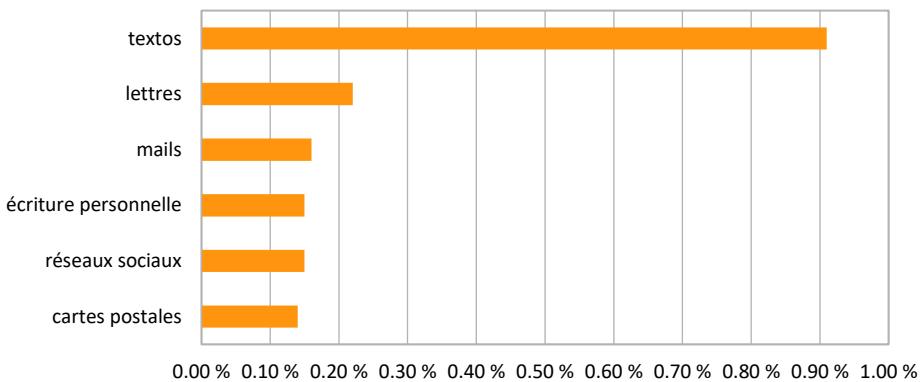
### 3. Résultats

Ces deux enquêtes font apparaître des tendances fortes communes aux collégiens et aux étudiants concernant la prégnance des écrits numériques et la reconnaissance de la valeur sociale de l'orthographe. Les contours de cette valeur sociale varient cependant entre les deux groupes du fait de leur différence d'âge et de la plus grande proximité des étudiants de STS avec le monde professionnel.

#### 3.1. Domination des écrits numériques dans les pratiques extrascolaires déclarées

Les discours recueillis auprès des élèves confirment que les outils numériques occupent une place de premier plan dans les pratiques d'écriture extrascolaires qu'ils déclarent.

Nous avons demandé aux élèves de troisième interrogés en 2016 ce qu'ils avaient l'habitude d'écrire. Le diagramme en barres ci-dessous représente les réponses que nous avons recueillies.



**Figure 2 - Pratiques citées par les élèves de troisième de 2016 (en pourcentage des élèves interrogés. N'ont été conservés que les valeurs supérieures à 10%)**

On remarque une grande diversité des pratiques d'écriture sous forme numérique (mails, réseaux sociaux) mais aussi sous forme manuscrite notamment dans le cadre de la correspondance ou de toutes les formes d'écriture personnelle. Néanmoins, la pratique la plus courante, et de loin, est l'écriture de textos puisque 90 % des élèves en écrivent.

Les réponses portées sur le questionnaire par les 178 élèves de STS interrogés en 2017 montrent une prééminence encore plus nette des écrits numériques chez ces étudiants âgés de quatre à cinq ans de plus que les élèves de troisième. Aucun élève n'a coché aucune case. L'écriture de textos est la seule pratique partagée par tous, suivie par l'écriture de mails et de messages sur réseaux sociaux. Pour autant, les supports cités sont assez diversifiés et l'écriture sur papier occupe encore une place non négligeable. À peu près le quart des élèves dit écrire des cartes postales et des lettres. Malgré cela, les pratiques les plus partagées par les élèves de nos deux corpus sont bien des pratiques numériques.

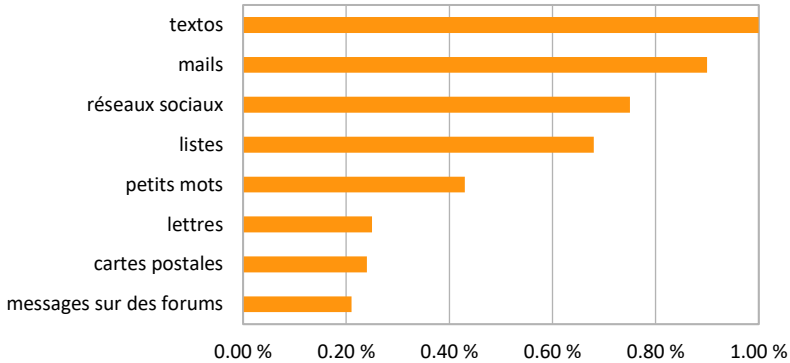


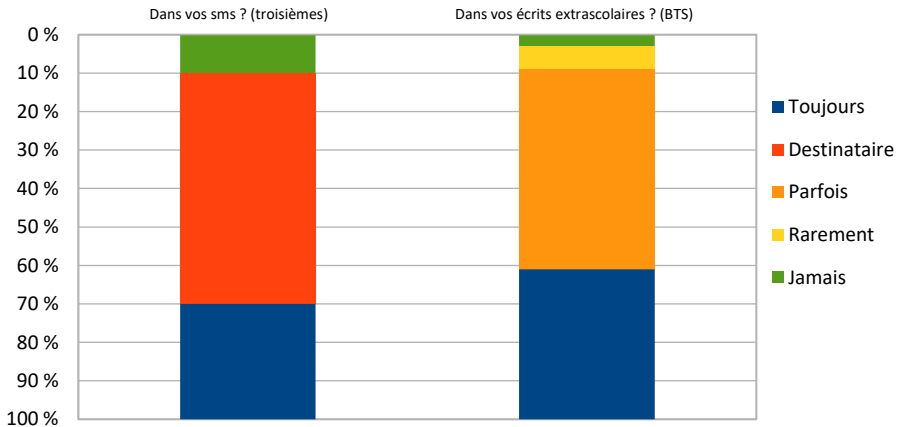
Figure 3 - Pratiques citées par les élèves de STS de 2017 (en pourcentage des élèves interrogés)

### 3.2. *Un trait commun : l'adaptation au destinataire et au contexte d'écriture*

Les deux enquêtes nous ont amenés à interroger les élèves sur le respect, ou non, de la norme orthographique lors de l'écriture hors contextes professionnel et scolaire. On a systématiquement demandé aux élèves de troisième s'ils respectaient l'orthographe dans leurs textos. Néanmoins, quelques élèves de troisième ne pratiquant pas l'écriture de SMS, nous ne disposons que de 116 réponses exploitables sur 131 entretiens. Celles-ci ont pu être réparties en trois catégories : *jamais*, *selon le destinataire* et *toujours*. Les 178 étudiants de STS ayant rempli le questionnaire ont dû indiquer s'ils respectaient *jamais*, *rarement*, *parfois* ou *toujours* l'orthographe dans leurs écrits hors contexte professionnel et scolaire. Les deux diagrammes ci-dessus représentent la répartition des réponses dans les deux groupes. Il ne s'agit pas exactement du même type de données, mais elles nous semblent comparables puisqu'elles représentent toutes l'importance accordée à l'orthographe en contexte privé d'après les déclarations des scripteurs.

La répartition des deux groupes d'enquêtés est comparable même si les élèves de STS répondent plus rarement qu'ils ne font jamais attention et plus souvent qu'ils y font toujours attention. Les 65 entretiens de STS permettent de confirmer que la réponse « parfois », ainsi que la réponse « rarement », correspondent dans la très grande majorité des cas à une adaptation au destinataire. Sur les 30 élèves ayant apporté une de ces réponses et l'ayant commentée de façon exploitable, 24 disent prendre en compte le destinataire. Sur les 6 élèves qui n'évoquent pas explicitement le destinataire, 4 mentionnent le type d'écrit (officiel ou lettre de motivation) et 2 le temps disponible pour écrire.

Ces données montrent que les variations diaphasiques, c'est-à-dire liées au moment et au



**Figure 3 - Réponse à la question "Faites vous attention à l'orthographe?"**

contexte dans lequel l'énoncé est formulé, mises en avant par Françoise Gadet (2003), se retrouvent ici à l'écrit en particulier dans les écrits numériques dont les usages correspondent à des contextes divers que nous allons détailler dans les deux points suivants.

### 3.3. *L'orthographe, forme graphique de la norme sociale*

Lorsqu'on s'intéresse aux explications que donnent les élèves au fait qu'ils fassent attention ou non à l'orthographe, il apparaît que les réponses positives sont liées au poids de certaines structures sociales.

#### 3.3.1. *L'importance de la famille dans la transmission de la norme*

Chez les élèves de troisième interrogés en 2016, la famille apparaît comme un vecteur important de la norme puisque de nombreux enquêtés déclarent faire plus attention lorsqu'ils s'adressent à leurs parents qu'à des amis (23) ou à des adultes qu'à des personnes de leur âge (15). Chez les STS de 2017, seuls cinq élèves sur 65 disent faire une différence entre les amis et la famille. Quatre de ces élèves font plus attention pour leur famille que pour leurs amis à l'image de ce qu'on observait en troisième, mais le cinquième fait plus attention pour ses amis chez qui il a constaté une attention à l'orthographe que pour sa famille qui, selon lui, ne s'y intéresse pas. Par ailleurs, deux élèves témoignent d'une pression familiale depuis l'enfance pour le respect de l'orthographe et ils font tous deux partie de ceux qui disent faire toujours attention à l'orthographe.

On peut supposer que la différence observée entre les deux groupes est liée à l'âge des enquêtés même si les deux populations sont par ailleurs différentes. En effet, les STS arrivent à l'âge adulte et ce n'est plus prioritairement la famille qui incarne le poids de la norme orthographique hors cadre scolaire.

### 3.3.2. *Le poids des pratiques professionnelles chez les STS*

La nécessité de l'orthographe en contexte professionnel apparaît déjà chez les élèves de troisième mais de façon très ponctuelle parce qu'ils n'ont majoritairement jamais été confrontés à un contexte professionnel.

Chez les STS, La quasi totalité des élèves interrogés en entretiens dit faire attention en contexte professionnel (63 élèves sur 65) lors des envois de mails ou de textos à des professionnels, mais aussi lorsqu'ils rédigent des CV et lettres de motivation qui passent aujourd'hui par voie numérique, que ce contexte professionnel soit vécu (stages voire réelle expérience professionnelle) ou anticipé. Les deux étudiants qui disent ne jamais faire attention à l'orthographe en contexte professionnel sont scolarisés en STS Systèmes d'Informations aux Organisations et n'avaient pas encore d'expérience de contacts écrits avec des professionnels de leur domaine contrairement à plusieurs autres élèves de la même filière.

Conformément au cadre d'une formation professionnalisante, le poids de l'orthographe en contexte professionnel est étroitement associé au poids de l'orthographe en contexte scolaire. L'une des sept classes dont les entretiens ont été transcrits (12 entretiens) a expliqué travailler l'orthographe avec son professeur de français sous forme de dictées. Tous les élèves évoquant cette expérience ont à la fois témoigné d'un manque d'enthousiasme vis-à-vis du fait de refaire des dictées et d'une adhésion à la nécessité de travailler l'orthographe pour leur avenir professionnel.

### 3.3.3. *Internet : média de référence de la norme orthographique*

Lors de l'entretien, nous avons demandé à certains élèves s'il leur arrivait d'utiliser des outils pour corriger leur orthographe. Les 49 élèves de STS qui ont répondu positivement citent des outils numériques même s'ils utilisent aussi parfois des dictionnaires papier, le *Bescherelle* ou leur entourage. Les mentions d'utilisation de correcteurs de façon volontaire sont minoritaires mais existent : 5 disent utiliser le correcteur du traitement de texte et 5 le correcteur du téléphone. Un élève dit même utiliser le correcteur du téléphone pour vérifier des mots quand il n'est pas sûr de leur orthographe en cours de français (tout en précisant que le téléphone est interdit conformément au règlement de l'établissement...).

La technique la plus citée est cependant l'utilisation d'internet suivant deux techniques différentes. 20 enquêtés déclarent utiliser simplement la fonction de recherche de *Google*. Ils se contentent de noter le mot ou la phrase dans la barre de recherche de *Google* et adoptent la forme proposée par le moteur de recherche qui fait alors indirectement fonction de correcteur orthographique<sup>3</sup>. 33 enquêtés disent quant à eux aller voir les sites proposés, souvent en comparant les formes ou explications fournies par plusieurs sites différents. 8 étudiants se montrent capables de citer un site précis : un dictionnaire en ligne (*Larousse* ou *L'Internaute*) ou les correcteurs des sites *Reverso* ou *BonPatron*.

Le fait que la majorité des étudiants dit recourir à des outils pour corriger son orthographe confirme l'importance qu'ils accordent à la correction de leurs écrits au moins dans certains contextes. Ces outils de correction sont aujourd'hui très majoritairement des outils numériques exploités avec plus ou moins de discernement.

## 3.4. *L'apparition de formes d'écriture spécifiques*

Ce qui domine dans les pratiques numériques des élèves, c'est donc le respect déclaré de conventions sociales qui ne sont pas spécifiques à ces supports d'écriture. Néanmoins,

---

<sup>3</sup>Ce qui est nécessairement une technique risquée puisqu'à partir du moment où une forme est attestée *Google* peut la proposer.

certaines données recueillies montrent qu'il existe des pratiques d'écriture dont le rapport à la norme est lié à des caractéristiques dépendantes du support qui les porte.

### 3.4.1. *Une pratique spécifique de l'abrégé qui ne correspond pas à un relâchement de la norme orthographique.*

Chez les élèves de troisième interrogés en 2016, il est souvent difficile de distinguer ce qui relève d'une pratique alternative assumée adaptée à un outil et ce qui relève, à leurs yeux, de la simple négligence orthographique.

Dans les discours recueillis auprès des élèves de STS en 2017, la différence entre pratique de l'abrégé et négligence orthographique est souvent clairement énoncée. 16 élèves sur 65 témoignent d'une pratique actuelle de l'abrégé et 4 d'une pratique passée à laquelle ils ont renoncé du fait de difficultés de communication avec leur entourage. Ceux qui abrègent disent utiliser cette façon d'écrire dans les textos pour aller plus vite. L'abrégé apparaît donc comme un procédé adapté à un contexte d'écriture. 7 de ces élèves soulignent d'ailleurs qu'ils raccourcissent les mots mais ne renoncent pas à la correction orthographique. L'extrait d'entretien ci-dessous témoigne de cette attention accordée à la norme malgré l'utilisation de l'abrégé.

Chercheuse : Quand c'est des amis c'est juste que vous abrégez les mots ou est-ce que vous faites pas attention aux accords...

Élève : Si, j'écris quand même correctement mais il y a des mots qui sont raccourcis.

Chercheuse : C'est juste pour aller plus vite.

Élève : Oui, voilà.

Chercheuse : Mais vous réfléchissez quand même aux accords, à écrire correctement.

Élève : Oui.

L'attention accordée à l'orthographe s'incarne ici dans le respect des accords. D'autres enquêtés disent supprimer des lettres mais conserver les lettres du mot en abrégeant par exemple « pourquoi » en « pcq » plutôt que « pck ». Les données recueillies auprès de ces jeunes scripteurs confirment donc que la pratique du texto n'implique pas un relâchement du respect de la norme orthographique. C'est plutôt le rapport à la norme des scripteurs qui influence la façon dont ils gèrent leurs textos.

### 3.4.2. *Des pratiques numériques originales en forte adéquation avec la norme*

Certaines pratiques numériques semblent renforcer ce rapport à la norme. Des pratiques d'écriture intensive intrinsèquement liées à internet apparaissent dans les entretiens avec des élèves de STS. Celles-ci sont à l'origine chez certains élèves d'une très forte adhésion à la norme orthographique.

Un témoignage souligne l'importance de respecter la norme pour être crédible dans une écriture sur les réseaux sociaux.

Chercheuse : Et vous savez à quel moment vous avez commencé comme ça à essayer d'améliorer votre orthographe ?

Élève : Oui c'est quand j'ai commencé à beaucoup écrire sur internet, c'était quand j'étais en quatrième.

Chercheuse : Et c'était lié à votre pratique d'écriture sur internet ?



Élève : Oui parce qu'on n'est pas pris au sérieux quand on écrit avec plein de fautes, juste un simple poste Facebook avec plein de fautes, on n'a plus aucune crédibilité. C'est juste pour ça du coup.

Il s'agit du seul entretien attribuant un tel rôle à un réseau social généraliste. D'autres se plaignent au contraire de la qualité orthographique de ce que les gens écrivent sur internet et témoignent d'interventions, parfois à mauvais escient, pour corriger l'orthographe des autres. Dans tous les cas, il s'agit bien de signaler que le respect ou l'irrespect de la norme orthographique joue un rôle dans leur expérience d'écriture sur les réseaux sociaux.

Cependant, ce rôle semble plus marqué sur des sites spécialisés. Un élève évoque ainsi son expérience d'un forum de poèmes.

C'est un forum où il y a que des poètes, enfin des personnes comme moi qui se font un compte, qui ont leurs poèmes et comme auteur, des commentaires etc. Après y a des modérateurs, qui m'envoyaient des tas de messages pour me dire attention vous avez autant de fautes d'orthographe, vous avez un malus de un point et (...) ils vous ferment le compte en fait.

Internet ouvre ici un nouveau mode de publication des textes qui n'existait pas auparavant. Chacun a la liberté de publier comme il l'entend à condition de respecter les conditions du site qui, dans un domaine lié à l'écriture, impliquent un certain degré de correction orthographique. Dans un même ordre d'idée, 3 élèves<sup>4</sup> signalent leur participation à des forums *role play*, c'est-à-dire à des jeux de rôles écrits sur internet dans un univers défini. Nous reproduisons ci-dessous trois extraits témoignant du rapport à la norme dans ces pratiques d'écriture.

Dernièrement j'ai trainé sur des forums *Role Play* où on écrivait beaucoup. Du coup je sais que des fois on me corrigeait beaucoup, du coup après, à force de se faire corriger, on prend des habitudes au niveau de l'écriture.

Oui, c'est pour rendre quelque chose immersif, il faut quand même que les phrases aient du sens, que les temps soient bien accordés, qu'on sache utiliser les temps du récit, les temps de description etc.

C'est souvent les corrections parce que quand on crée ses personnages on doit faire une fiche pour expliquer le contexte du personnage, tout ça, son histoire, son caractère. Et en gros des fois les gens, les modérateurs, les administrateurs, ils nous font remarquer nos fautes pour éviter qu'on les reproduise dans les *RP*.

Ces trois extraits révèlent une pratique d'écriture collective qui exige le respect des normes de l'univers mais aussi de la langue pour des commodités de lecture et parce que ces communautés y semblent fortement attachées. Ce respect de la norme apparaît d'ailleurs autant comme un respect stylistique qu'orthographique. Cette pratique est très minoritaire mais pas isolée même si elle semble plus liée à des milieux technophiles. En effet, elle découle des possibilités ouvertes par internet et l'écriture en réseau.

En guise de conclusion, voici une citation d'un élève qui expliquait que l'orthographe était particulièrement importante dans le cadre professionnel du fait de la place grandissante des écrits numériques.

À notre époque, on communique beaucoup par internet, donc c'est un peu ce qu'on renvoie, notre orthographe, c'est un peu l'image qu'on renvoie, c'est

---

<sup>45</sup> élèves déclarent une pratique de ce type dans les questionnaires dont 4 élèves de STS Systèmes d'information aux organisations et une élève de STS assistant de manager.

comme si on était bien... avoir une bonne orthographe, c'est comme être bien habillé dans la vraie vie.

Conformément à ce qu'expliquent Fayol et Jaffré (2014), les élèves interrogés, et en particulier les plus âgés et les plus proches du monde professionnel, témoignent d'une forte adhésion à la norme orthographique. L'usage d'outils numériques semble majoritairement renforcer, plutôt que diminuer, l'importance de cette norme dans leurs écrits. Ceci ne signifie pas qu'ils écrivent partout de la même manière : en diversifiant les contextes de socialisation écrite, les écrits numériques diversifient également les normes en vigueur. Les données exploitées ici permettent d'effleurer cette question, mais la masse de données présentes sur internet en fait un terrain très prometteur pour les linguistes de l'écrit.

### ***Références bibliographiques***

- FAYOL, M. & JAFFRE, J.-P. (2014). *L'orthographe*. Paris: Presses universitaires de France.
- GADET, F. (2003). *La variation sociale en français*. Gap Paris: Ophrys.
- JOANNIDES, R. (2014). *L'écriture électronique des collégiens : quelles questions pour la didactique du français ?* Thèse de doctorat en sciences du langage. Université de Rouen.
- MANESSE, D., COGIS, D. & DORGANS-ROBINEAU, M. (2007). *Orthographe : à qui la faute?* Issy-les-Moulineaux (Hauts-de-Seine): ESF éditions.
- MILLET, A., LUCCI, V. & BILLIEZ, J. (1990). *Orthographe mon amour*. Grenoble: Presses universitaires de Grenoble.
- MORTAMET, C. dir. (2019). *L'orthographe : pratiques d'élèves, pratiques d'enseignants, représentations*. Rouen : Presses Universitaires de Rouen.
- PAVEAU, M.-A. & ROSIER, L. (2008). *La langue française passions et polémiques*. Paris : Vuibert.

### ***Références aux sites Internet***

- ANDREU, S. & STEINMETZ, C. (2016). *Les performances en orthographe des élèves en fin d'école primaire (1987-2007-2015)*. <http://www.education.gouv.fr/cid23433/les-performances-en-orthographe-des-eleves-en-fin-d-ecole-primaire-1987-2007-2015.html>, consulté le 12/04/2018.
- BERNICOT, J. (2013). *La pratique des SMS des collégiens et des lycéens*. <http://www.ires.fr/index.php/etudes-recherches-ouvrages/etudes-des-organisations-syndicales/item/4884-la-pratique-des-sms-des-collegiens-et-des-lyceens>, consulté le 11/04/2017.

## **UNDERSTANDING, SETTING UP AND EVALUATING A BLENDED LANGUAGE COURSE: A SYSTEMIC APPROACH**

Aude LABETOULLE<sup>a</sup>

[aude.labetouille@lecnam.net](mailto:aude.labetouille@lecnam.net)

<sup>a</sup>*UR Formation et apprentissages professionnels, Cnam, Paris, France*

---

### **1. Introduction**

Blended learning has become increasingly frequent at the university in the past 20 years in higher education, and language learning is no exception (Nissen, 2014: §2). Literature concerning blended language learning is now extensive (see for example Neumeier, 2005, Nicolson *et al.*, 2011, Farr & Murray, 2016, McCarthy, 2016) and provides for a wide range of analyses. Definitions of blended learning often include “a mix of face-to-face and distance learning, with both elements being an important part of the whole” (Sharma & Westbrook, 2016: 320), as well as e-learning. It is also often highlighted that blending a course considerably complexifies the learning and teaching processes (Garrison & Kanuka, 2004: 96, Nissen, 2019: 14-17). Blended language learning is specific in that because language is both the objective and the means of teaching and learning, language blended courses are often characterized by active participants, strong interactions, a great diversity of resources and is often set in the tradition of “learning to learn” (Nissen, 2019: 12).

The present contribution consists in shedding a new light on blended learning thanks to a systemic approach. We argue this approach is particularly appropriate to understand complex research objects such as blended language learning training courses. We aim to demonstrate that it was instrumental in all the steps involved in an action-research, in setting up and evaluating an English training course at the University of Lille (France) for undergraduate musicology students<sup>1</sup>.

With this aim in view, we first define what is here called a systemic approach. Next, we present the way this approach was used to analyse the learning environment in which the new training course was to take place. We then describe the new blended training course which was designed based on the results of the needs analysis. To finish, we evaluate the blended system and discuss to what extent the systemic approach was relevant to understand this aspect of the training course.

### **2. Definition of the blended language training course with a systemic approach**

Setting up and evaluating a training course is no easy task and despite scientific progress, teachers are still left struggling with an intimidating number of questions. How should the learning context be analysed? How can working conditions, competing views – of learners, teachers, supervisors, administrative staff – on objectives, contents, learning and teaching methods be all taken into account? What should the objectives of the course be? How can learner motivation and autonomy be both encouraged? And most importantly, how can all these various elements be articulated? In view of how complex analysing a learning environment and setting up a course is, one of the main criteria for the selection of the

---

<sup>1</sup> In France, studying a second language is compulsory in most higher education programmes. LANSOD stands for “LANguage for Students of Other Disciplines” and refers to language classes destined to students whose major is not languages, but other disciplines such as medicine, chemistry, etc. 90% of students who have language classes in higher education take LANSOD courses (Causa & Derivry-Plard, 2013).

theoretical framework in which the action-research was to be grounded was that it should take complexity into account.

Systemic theories are closed to complexity theories such as those developed by Edgar Morin (1990) and Jean-Louis Le Moigne (2012). They are regularly adopted in studies of language learning and acquisition (Bertin, Gravé & Narcy-Combes, 2010, De Bot, Lowie & Verspoor, 2007, Larsen-Freeman and Cameron, 2008, Rivens Mompean, 2013, Waninge, De Bot & Dörnyei, 2014). Schematically, systemics invites us to view training courses as systems; they are made up of elements which interact with one another to form a whole. Changing one element in the system will impact the whole system. Here is one example:

The components of the system continually interact and are mutually dependent.  
 Illustration: a change in the curriculum entails pedagogic adjustments in terms of contents, methods and modes of teaching/learning. The other components are affected by these changes. (Bertin *et al.*, 2010: 5)

Likewise, choosing to adopt a blended format will necessarily impact learners, teachers, contents and the learning and teaching processes.

Thus, the systemic approach first provides us with an ontological characterization of blended learning training courses. Let us now have a look at how this approach guided the analysis of the learning environment in which a new training course was to be set up.

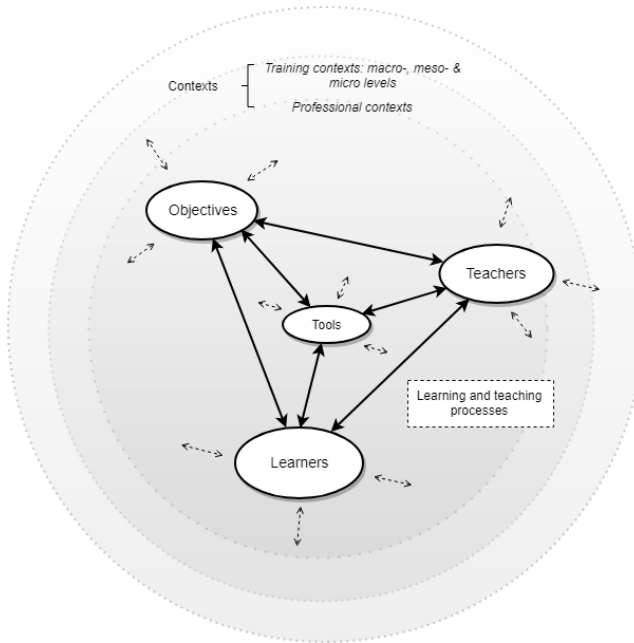
### **3. Analysis of the learning environment**

#### **3.1. Design of a model of learning environment**

To structure the analysis of the learning environment, we first modeled a generic learning environment, as represented in Figure 1 below. Modelling is a common tool in systemic approaches (Rosnay, 1975: 121). In our case, it provides for a global understanding of a complex system and it helps bring the relationships between the elements of a system to light (Durand, 2013: Chapter 3, §36-37).

As it is impossible to identify all the elements and processes involved in a learning environment, we selected the elements to appear on the model with the concept of “simplicity” in mind. According to Berthoz (2018), it is “simplifying principles” which “enable the brain to prepare an action and plan for the consequences of it”, “despite the complexity of natural processes” (back cover).

The model is based on an analysis of didactic models (Bertin *et al.*, 2010, Carré, Moisan & Poisson, 1997, Houssaye, 1988, Legendre, 1988 and Germain, 1989, Rézeau, 2001, Rivens Mompean, 2013). It is in keeping with our systemics approach in that it models the learning environment as made up of interacting systems which are mutually dependent. There are five main elements: learners, teachers, objectives of the course, tools and contexts. The context is multi-layered; it ranges from European language policies and the local language policy of the school/university, to the premises and the equipment, as well as the academic and professional context – in our case study, the musicology department and the students’ future occupations. The tools play a mediation role between content and learners and teachers – for example, course materials, the syllabus, as well as digital tools. The main processes at work in the system are the learning and teaching processes.



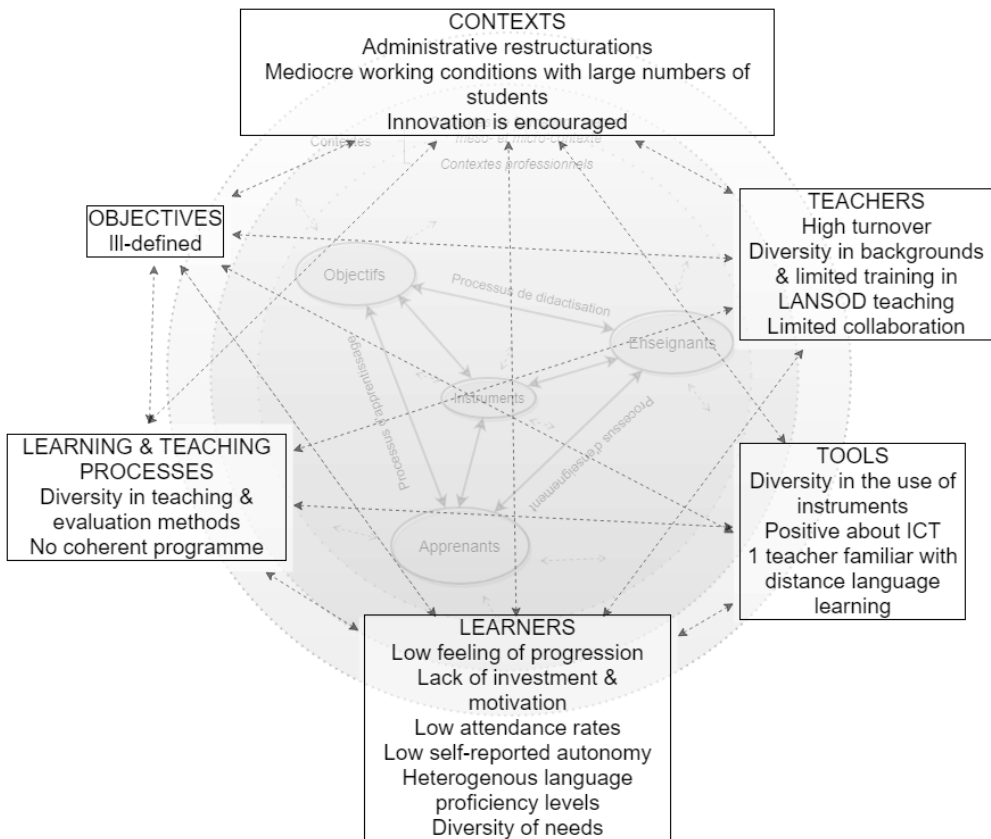
**Figure 4 – Model of a generic learning environment**

### ***3.2. Needs analysis and the objectives of the new course***

The next step was a needs analysis. It consisted in using this generic model to guide the analysis of the specific context of the English training course for undergraduate musicology students at the University of Lille. To do so, questionnaires were submitted to the learners and former undergraduate musicology students, as well as the English teachers and the supervisors. Two facts stood out:

1. When asked whether they were satisfied with the classes they had so far, the 43 learners who responded rated their English classes 2.3 (out of 5) on average, with clear differences between first year (Y1) students (3.3) and Y2 students (1.5), when Y3 students gave the course an average rating of 2.5. The scores given by the four teachers who responded varied a lot, ranging from 1 to 5.
2. Despite there being 12 two-hour English classes for each of the 6 semesters of the undergraduate programme, 58% of the learners declared that they felt they had stagnated or regressed since starting university.

The data gathered from the needs analysis, summarized in the figure below (Figure 2), helps explain these unsatisfactory results:



**Figure 2 – Description of the learning environment**

Data suggests the objectives of the course were ill-defined and that there was no coherent programme from Year 1 to Year 3. The working conditions were mediocre in that there were many students per group (between 30 and 50) and technological equipment was often lacking in classrooms. The learners had heterogeneous language proficiency levels (ranging from A1 to native speakers) and different needs in English, as their professional objectives varied (professional musician, primary school teacher, music teacher in secondary school amongst others). Half of the students declared they were not autonomous language learners. A majority of them also stressed they lacked motivation, especially as they were often quite engaged in extra-curricular activities (such as giving music lessons). In this university attendance is not compulsory (except for evaluations) and absenteeism rates were quite high as most teachers indicated hardly ever having more than half of the students in Year 2 and Year 3. Teachers rarely taught for more than a semester and they had little if no training in LANSOD teaching. Collaborative work was encouraged by the coordinator but the attempts were seen as timid and not always successful.

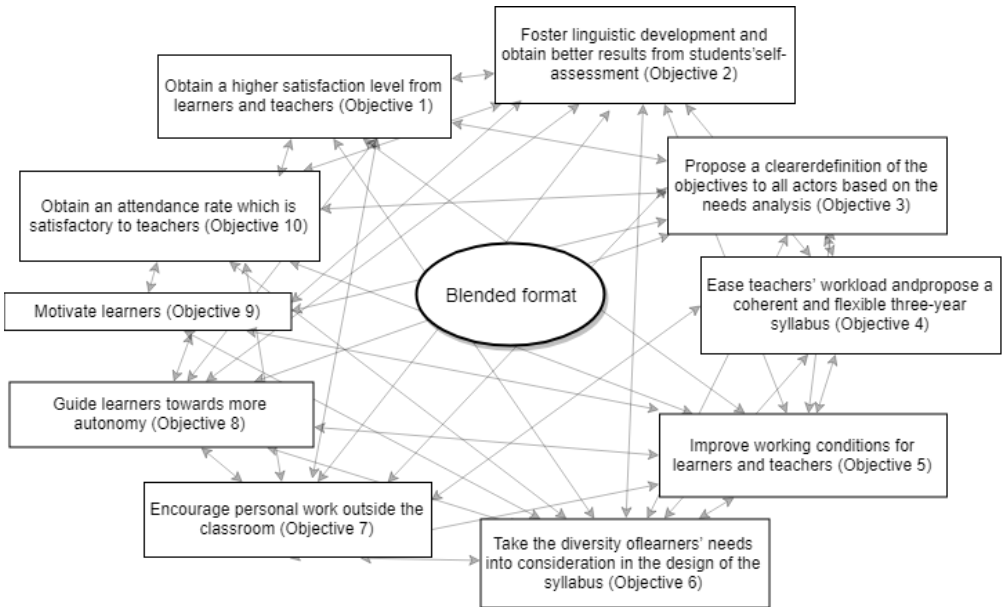
Based on these elements, ten objectives were selected for the new training course: obtaining a higher satisfaction level from learners and teachers (Objective 1), obtaining better results from students' self-assessment of linguistic development (Objective 2), proposing a clearer definition of the objectives to all actors (Objective 3), reducing teachers' workload and propose a coherent and flexible three-year syllabus (Objective 4), improving working conditions for learners and teachers (Objective 5), taking the diversity of learners' needs into consideration in the design of the syllabus (Objective 6), encouraging investment and personal

work outside the classroom (Objective 7), guiding learners towards more autonomy (Objective 8), motivating learners (Objective 9), as well as obtaining an attendance rate which is satisfactory to teachers (Objective 10).

## 4. Presentation of the blended course

### 4.1. Rationale for a blended format

The systemic conception of a training course led us to state that a blended format would necessarily impact learners, teachers, contents and the learning and teaching processes. Therefore, when setting up the course, we hypothesized that choosing a blended format would help us reach all the objectives stated above, as it is underlined in Figure 3 below:



**Figure 3 – Potential impact of the blended format on the objectives of the course**

Let us see precisely how we intended to reach the ten goals thanks to a blended approach. The blended format was aimed to help reach Objectives 5 and 6, that is improving working conditions of learners and teachers and taking the diversity of learners' needs into consideration in the design of the syllabus. Because there were many students per group and the learners displayed heterogenous levels, the blended system would enable us to divide groups into smaller groups based on language proficiency. Even though the learning objectives would be the same for both advanced and intermediate students, there could be several ways to differentiate contents (Perrenoud, 2005: 29): in addition to a core syllabus, specific activities could be provided for both groups both in class and online (Nissen, 2019: 48), and the in-class pace could be adjusted accordingly. A blended system would also provide for a variety of learning approaches and materials, and a blended system would provide the learners with more flexibility in terms of time, place and pace (for example Nicolson *et al.*, 2011, Sharma & Westbrook, 2016).

To encourage investment and personal work outside the classroom (Objective 7), the online modules could be closely integrated with the face-to-face classes in a flipped classroom approach; content delivery seen online would have to be mastered to perform tasks in class. The online modules could be marked to foster motivation and thus to encourage students to be exposed to English at least once every two weeks if they did not come to face-to-face classes.

As completing an online module could be considered as attendance, the blended approach might help improve the attendance rate overall and hence reach Objective 10.

A blended course necessarily implies greater autonomy from learners who have to adapt to an unusual format (Objective 8). As Linda Murphy and Margaret Hurd underline,

A blended learning environment provides more/different choices and decisions for learners, so arguably increases the opportunity for learners to exercise their capacity for autonomy. (2011: 45)

Teacher support is considered paramount, so each module could end on a feedback activity to be discussed in face-to-face classes.

The objectives and the way the course works, the description of each face-to-face class as well as the calendar of the course could all be presented on the online learning platform so as to propose a clearer definition of the objectives to all actors based on the needs analysis (Objective 3). Because of the high turnover of the teacher squad, setting up a blended system might help define more coherent objectives for the three-year undergraduate programme. It could be designed so that teachers would have a comparable workload with face-to-face only classes (Objective 4).

Choosing a blended format is often justified as a means to optimize the language learning process, that is to reach Objective 2 here. However, the arguments are generally not compelling enough:

[There is] a high degree of agreement between the reasons given in the literature for blending and the reasons given by language providers in the case studies they describe, in particular with regard to increased access, flexibility and “improved pedagogy” (Wittaker & Tomlinson, 2013). The latter is often cited, but it is difficult to pin down what is meant and it is generally left unsupported. (Sharma & Westbrook, 2016: 324)

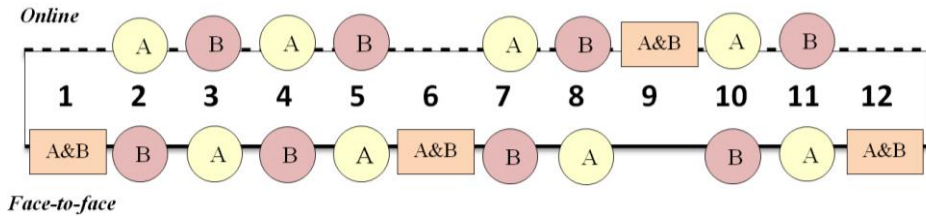
In our case, we hypothesized that thanks to a close integration of the online modules and the face-to-face classes with a flipped classroom approach and more time spent on production in class, learners would improve. Likewise, our systemic stance led us to hypothesize that by trying to reach the other objectives, thanks to the blended format especially, the new course would foster learner motivation (Objective 9) as well as teacher and learner satisfaction (Objective 1).

#### **4.2. *General presentation of the blended course***

The new training course followed these guidelines. In concrete terms, in semester 1, the course combined face-to-face and online sessions, group (A&B) and half-group (A/B) sessions as illustrated in Figure 4 below. The online modules were specifically designed for this course on the Moodle learning platform and were mostly devoted to receptive skills, when face-to-face classes were more focussed on oral and written production.

These groups then alternated between face-to-face and online sessions. For example, in week 1 all the students came to class; in week 2 group A (the advanced group) did the online module and then came to class in week 3, while group B (the lower proficiency group) came to class on week 2 and did the online module in week 3.





**Figure 4 – Organisation of the blended course (semester 1)**

The syllabus was organized around tasks. These tasks were mostly focussed on contents linked to music, with a gradual transition towards more professional English from Y1 to Y3. Learners were encouraged to work on the instructional content online and then perform tasks in class, such as “pitching the project of a music festival to sponsors,” “designing the programme of the festival” in Year 1, “teaching an introductory music class” and “design a professional resume” in Year 3.

## **5. Impact of the blended format on the course**

### **5.1. Results**

The new blended training course was then set up over the course of four semesters, from September 2016 to April 2018. At the end of each semester, the course was evaluated thanks to questionnaires submitted to the learners and the teachers, teachers’ weekly diaries and data obtained on Moodle<sup>2</sup>. The results detailed below concern the first semester.

Based on the data collected, it appears most of the objectives were reached. Learners and teachers alike could explicit the main objectives of the course (Objective 3). Thanks to the blended system, we were able to divide each group in two level groups. Learners and teachers alike scored the groups based on language proficiency 4.4 out of 5. The teachers also remarked there had been more teamwork. Overall, the working conditions of learners and teachers thus improved (Objective 5). 56% of learners were satisfied with their investment in the course; when 20% had previously declared working regularly, the number rose to 60% with the new course (Objective 7). 67% of learners wrote they were now more autonomous language learners (Objective 8). 69% of learners declared having come to all or all but one face-to-face class. 71% of the students completed all the online modules (Objective 10).

Choosing a blended format thus helped us reach the main objectives of the course, that is get a higher satisfaction rate from learners and teachers (Objective 1) and obtain better results from students’ self-assessment (Objective 2). Indeed, the learner’s satisfaction rate went from 2.3 to 3.7. Teachers’ scores were more homogenous than before and also improved to 3.7. 73% of learners considered they had progressed. The great majority of learners considered the training course was adapted to their needs as 71% scored it 4 or more out of 5 (Objective 6). Learners also declared being more motivated than before, assessing their motivation levels during the semester at 3.6 on average (Objective 9).

However, the workload of the teacher in charge of creating the course was considerable (Objective 4).

## **6. Discussion**

The new training course seems to have been successful, and the blended approach appears to have been instrumental in its success. However, the blended system is not the only element which accounts for these results and the analysis above should be read with caution.

<sup>2</sup> Thus, the evaluation relied on qualitative methods essentially, as we aimed at collecting the actors’ representations of the training course.

In a systemic approach, it is accepted that all elements and processes are linked and impact one another. Although in terms of epistemological and theoretical stance it helps us understand a training course better, this approach is nonetheless not methodologically very handy. Joël de Rosnay states that ideally, each element is considered individually and its influence on the behaviour of the various sub-components of the system is evaluated (1975: 123). Some studies of restricted phenomena in second language learning adopting this methodology are very convincing. To investigate second language learner motivation in class, Freerkien Waninge, Kees De Bot and Zoltán Dörnyei (2014) carried out an individual-level microanalysis on four students by having the learners write down their levels of motivation every five minutes while researchers took many variables into account. In our study of a whole training course however, it becomes impossible to identify all factors that hindered or favoured a specific process in such a complex dynamic system. We are far from an experimental approach with which it is possible to isolate variables and assess their impact with a fairly high degree of certainty. One can only voice hypotheses by highlighting which elements and processes seem to have had the most significant influence on the training course, and only tentatively discuss the impact one element (here the blended system) seems to have had on the rest of the system.

Other factors account for the success of the course. Table 1 below is part of the evaluation board of the new blended course. It is a synthesis of the main problems that were identified thanks to the needs analysis, the objectives the new course aimed to meet, and the means to reach those goals.

Understanding, setting up and evaluating a blended language course: a systemic approach 34

Problems	Objectives of the new training course	Means to reach these objectives
Teachers' and learners' low satisfaction levels	1. Obtain a higher satisfaction level from learners and teachers	Take the results of the needs analysis into account to design the new training course
Learners' feelings of low progress	2. Foster linguistic development and obtain better results from students' self-assessment	Support motivation, investment and language practice
Ill-defined objectives	3. Propose a clearer definition of the objectives to all actors based on the needs analysis	Propose & set up a flexible syllabus based on the results of the needs analysis
Important turnover, little experience in LANSOD teaching for some, mediocre collaborative work	4. Ease teachers' workload and propose a coherent and flexible three-year syllabus	<ul style="list-style-type: none"> <li>- Propose &amp; set up a flexible syllabus based on the results of the needs analysis</li> <li>- Propose reusable teaching materials</li> <li>- Propose a training course encouraging collaboration</li> </ul>
Mediocre working conditions	5. Improve working conditions for learners and teachers	<ul style="list-style-type: none"> <li>- Reduce the number of students per group</li> <li>- Divide groups based on language proficiency</li> </ul>
Diversity of needs	6. Take the diversity of learners' needs into consideration in the design of the syllabus	<ul style="list-style-type: none"> <li>- Reduce the number of students per group</li> <li>- Divide groups based on language proficiency</li> <li>- Propose activities that foster individualized learning</li> </ul>
Little student investment	7. Encourage personal work outside the classroom	Ambitious and motivating syllabus
Learners' feelings of low language learning autonomy	8. Guide learners towards more autonomy	Syllabus and <b>teaching methods</b> aimed at fostering autonomous learning
Learners' unsatisfactory motivation levels	9. Motivate learners	<ul style="list-style-type: none"> <li>- Implement the objectives stated above</li> <li>- Reinstate the importance of language learning</li> <li>- Frequent evaluations</li> </ul>
Unsatisfactory attendance rate	10. Obtain an attendance rate which is satisfactory to teachers	<ul style="list-style-type: none"> <li>- Implement the objectives stated above</li> <li>- Make attendance more constraining</li> </ul>

**Table 2 – Evaluation board of the new English blended learning course: Problems, Objectives and Means**

The elements that are highlighted are considered to be directly linked to the blended dimension of the course. We can see how many other factors interact with this aspect of the course, which were also key in its success. For example, the learning objectives were better defined, the contents were more relevant to the learners' needs, and the teachers were highly invested.

## 7. Conclusion

The aim of this paper was to demonstrate how systemics can be instrumental in all the steps involved in setting up and evaluating a blended language training course. It helped us define the research object better, and it guided the needs analysis and the interpretation of the results. Indeed, the systemic approach first prompted us to define a training course as a complex system made up of interrelated sub-systems, elements and processes. This led to the creation of a model used as a grid for the needs analysis. The results of the needs analysis helped us have a global understanding of the various elements and processes at work in the learning environment, which in turn determined the design of the new training course. As far as interpreting the results is concerned, the principles on which the systemic approach relies led us to qualify the impact the blended system might have had on the overall positive results of the new training course, as other factors were also key in its success. In this respect, further analyses are required, by investigating more specifically how each pole of the model was impacted by the blended system for example.

## Bibliography

- BERTHOZ, Alain. (2009). *La Simplexité*. Paris: Odile Jacob.
- BERTIN, Jean-Claude, GRAVE, Patrick & NARCY-COMBES, Jean-Paul (2010). *Second Language Distance Learning and Teaching: Theoretical Perspectives and Didactic Ergonomics*. Hershey, PA: IGI Global.
- CARRE, Philippe, MOISAN, André & POISSON, Daniel (1997). *L'autoformation*. Paris: Puf.
- CAUSA, M., & DERIVRY-PLARD, M. (2013). Un paradoxe de l'enseignement des langues dans le supérieur : diversification des cours pour les étudiants et absence de formation appropriée pour les enseignants. In *Enseigner les langues à l'université au 21e siècle*. Paris: Riveneuve Éditions.
- DE BOT, Kees, LOWIE, Wander & VERSPOOR, Marjolyn (2005). *Second language acquisition: An advanced resource book*. New York: Routledge.
- DE ROSNAY, Joël (1975). *Le macroscope : vers une vision globale*. Paris: Éditions du Seuil.
- DURAND, Daniel. (2013). *La systémique*. Paris: Presses Universitaires de France.
- FARR, Fiona, & MURRAY, Liam. (2016). *The Routledge handbook of language learning and technology*. New York : Routledge.
- GARRISON, Randy & KANUKA, Heather. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7, 2: 95-105.
- GERMAIN, Claude (1989). Un cadre conceptuel pour la didactique des langues, *Études de linguistique appliquée*, 75: 61-77.
- HOUSSAYE, Jean (1988). *Théorie et pratiques de l'éducation scolaire*, « Le triangle pédagogique ». Berne: Peter Lang.
- LARSEN-FREEMAN, Diane & CAMERON, Lynne (2008). *Complex systems and applied linguistics*. Oxford: Oxford University Press.
- LE MOIGNE, Jean-Louis (2012). *Les épistémologies constructivistes*. Paris: Presses universitaires de France.
- LEGENDRE, Renald (1988). *Dictionnaire actuel de l'éducation*. Paris & Montréal: Larousse.
- MCCARTHY, Michael (2016). *The Cambridge guide to blended learning for language teaching*. Cambridge : Cambridge University Press.
- MORIN, Edgar (1990). *Introduction à la complexité*. Paris: Le Seuil.

- MURPHY, Linda & HURD, Stella (2011). Fostering learner autonomy and motivation in blended teaching, in *Language Teaching in Blended Contexts*, Nicolson, Margaret, Murphy, Linda & Southgate, Margaret (Eds.). Edinburgh, U.K.: Dunedin Academic Press, 43-56.
- NEUMEIER, Petra (2005). A closer look at blended learning - parameters for designing a blended learning environment for language teaching and learning. *ReCALL*, 17, 2: 163-178.
- NICOLSON, Margaret, MURPHY, Linda & SOUTHGATE, Margaret (2011). *Language Teaching in Blended Contexts*. Edinburgh, U.K: Dunedin Academic Press.
- NISSEN, Elke. (2014). Les spécificités des formations hybrides en langues. *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication*, 17.
- NISSEN, Elke (2019). *Formation hybride en langues : articuler présentiel et distanciel*. Paris: Didier.
- PERRENOUD, Philippe (2005, 3rd edition). *La pédagogie à l'école des différences, Fragments d'une société de l'échec*. Paris: ESF.
- REZEAU, Joseph (2001). « Médiation et médiatisation dans l'enseignement des langues en environnement multimédia ; le cas de l'apprentissage de l'anglais en DEUG d'Histoire de l'art à l'université ». Doctorat en didactique de la langue, Université Victor Segalen Bordeaux.
- RIVENS MOMPEAN, Annick (2013). *Le Centre de Ressources en Langues : vers la modélisation du dispositif d'apprentissage*. Villeneuve d'Ascq : Presses Univ. Septentrion.
- SHARMA, Peter, & WESTBROOK, Kevin (2016). Online and blended language learning, in *The Routledge handbook of language learning and technology*, Farr Fiona & Murray Liam (Eds.). New York: Routledge, 320-333.
- WANINGE, Freerkien, DÖRNYEI, Zoltán & DE BOT, Kees (2014). Motivational dynamics in language learning: Change, stability, and context, *The Modern Language Journal*, 98, 3.

# CORPUS D'APPRENANTS ET RECHERCHE LINGUISTIQUE. UNE ÉTUDE SUR L'ITALIEN.

Emanuele CASANI<sup>a</sup>

[emanuele.casani@unive.it](mailto:emanuele.casani@unive.it)

<sup>a</sup>*Dipartimento di Studi Linguistici e Culturali Comparati, Università Ca' Foscari Venezia  
Ca' Bembo, 30123 Venezia, Italia*

---

## 1. Introduction

Dès le lendemain de la publication du Cadre Européen Commun de Référence pour les Langues (CECRL), le Conseil de l'Europe a encouragé l'utilisation de corpus de langue authentique pour le développement d'outils d'aide au Cadre décrivant les compétences linguistiques et communicatives dans les différentes langues européennes (Council of Europe, 2005 ; Hulstijn, 2007). Toutefois, les réponses à cet appel ont été peu nombreuses et limitées aux langues européennes largement utilisées, à l'exception du *Norsk Profil* (Carlsen, 2013) norvégien. Pour la langue italienne, ces réponses se sont matérialisées par la publication du *Profilo della Lingua Italiana* (Spinelli & Parizzi, 2010) et, plus récemment, du corpus *MERLIN*<sup>1</sup>, qui est utilisé dans l'étude décrite ici (Casani, 2020). Cette étude illustre une application possible de la méthodologie d'analyse d'erreurs à un corpus d'apprenants (CA) italien. Le but est de décrire la compétence morphosyntaxique d'un échantillon représentatif d'apprenants adultes d'italien Langue Étrangère (LE) pour les niveaux élémentaires et intermédiaires du CECRL.

### 1.1. CA et recherche linguistique

Dès le début des années 1990, la linguistique acquisitionnelle et la linguistique de corpus ont fait converger leurs études dans la recherche sur les CA (Leech, 1998 ; Tan, 2005), vastes et systématiques recueils électroniques de productions d'apprenants de langue seconde (L2) et LE (Granger, 2002 ; Leech, 1998).

De la linguistique de corpus, la recherche sur les corpus d'apprenants a adopté l'approche quantitative des mécanismes d'analyse, tels que l'annotation. De la linguistique acquisitionnelle, elle a hérité les méthodes d'analyse contrastive et d'analyse d'erreurs qui, mises en œuvre avec les possibilités de la recherche informatisée sur les corpus, ont généré un outil puissant pour l'étude quantitative et qualitative de l'apprentissage des langues. Travailler sur un corpus d'apprenants permet généralement de mettre en évidence les domaines dans lesquels les apprenants manifestent une sous-utilisation ou un abus des fonctions langagières dans une langue par rapport aux locuteurs natifs grâce à la méthodologie contrastive, et d'acquérir des connaissances sur les erreurs commises par les étudiants d'une certaine langue grâce à la méthodologie d'analyse d'erreur (Leech, 1998, Granger, 2002). Dans la première approche, il est possible de travailler directement sur les corpus ; alors que dans la seconde, un type spécifique de codage et d'annotation de l'erreur est requis. Pour cette raison, les CA sont particulièrement utiles quand ils sont étiquetés, c'est-à-dire quand toutes les erreurs du corpus ont été annotées avec un système normalisé de classification des erreurs, qui enrichit les données d'informations linguistiques. Les CA diffèrent des collections de données couramment utilisées par les chercheurs en linguistique acquisitionnelle et en didactique des langues pour deux raisons principales: ils sont informatisés et peuvent donc être analysés à l'aide d'un large éventail d'outils logiciels linguistiques assurant un traitement rapide et efficace des données; ils sont très vastes et fournissent donc une base beaucoup plus fiable pour la

---

<sup>1</sup> <https://merlin-platform.eu>.

description de la langue des apprenants que les collections utilisées précédemment (Granger, 2003). Si un corpus de 200 000 mots peut être considéré déjà très vaste pour la linguistique acquisitionnelle, des corpus de plusieurs centaines de millions de mots sont désormais devenus la norme (Granger, 2003).

## 1.2. Classifications des CA

La conception d'un CA peut se dérouler sur la base de critères linguistiques (mode, médium, genre ou sujet des textes); critères liés aux tâches (longitudinales/transversales, spontanées/préparées), critères relatifs aux apprenants (par exemple L2 ou LE, âge, sexe, langue maternelle, expérience à l'étranger etc.) (Tono, 2003). Certains critères de classification courants sont :

- la langue cible: la plupart des CA se réfèrent à l'anglais L2/LE. Le nombre de CA pour les autres langues est encore limité mais en augmentation (voir figure 1) ;
- le médium: les CA peuvent être composés de textes écrits ou parlés. Ces derniers sont beaucoup plus difficiles à remplir et donc moins communs ;
- langue maternelle : les données peuvent provenir d'apprenants de langue maternelle identique ou différente ;
- compétence dans la langue cible: certains CA rassemblent des textes d'apprenants du même niveau, d'autres incluent des textes de différents niveaux ;
- critères d'annotation: de nombreux CA ne contiennent que des données brutes, éventuellement corrigées, sans annotation linguistique; certains incluent l'annotation des parties du discours (Hana *et al.*, 2010).

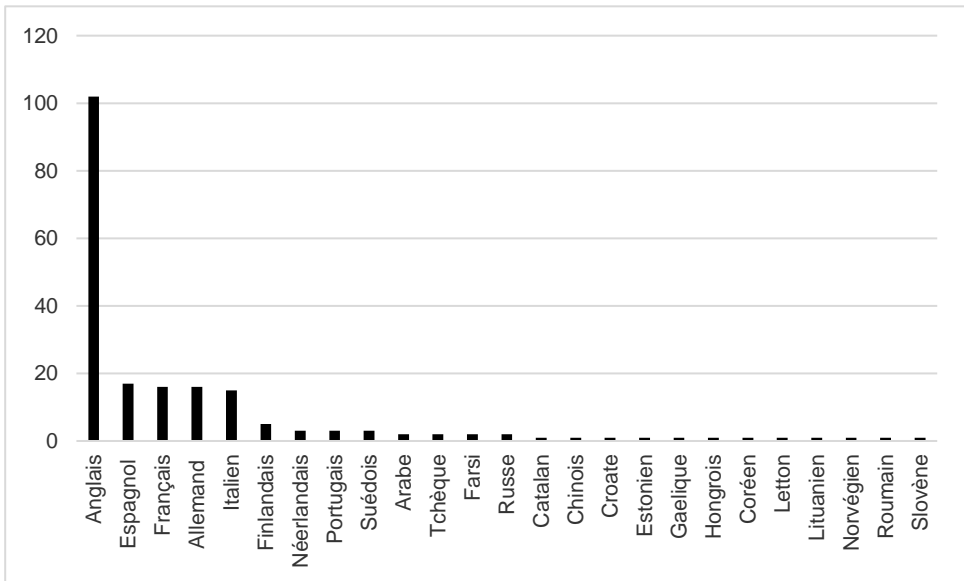


Figure 1 – CA existants classés selon la langue cible (Université Catholique de Louvain, 2018)

## 1.3. CA italiens

Comme on le voit dans la figure 1, 52% du total de 198 corpus existants<sup>2</sup> est dédié à l'anglais, suivi à grande distance de 9% de corpus espagnols et de 8% des corpus français et allemands.

<sup>2</sup> Données de l'Université Catholique de Louvain.

L'Italien est en quatrième position (7%) avec un nombre de 15 corpus dont 7 comprennent uniquement l'italien et les autres sont multilingues<sup>3</sup>.

Malgré le vaste corpus de recherches sur l'acquisition de la L2, l'italien est actuellement largement sous-représenté en termes d'application de méthodologies quantitatives et statistiques aux données des apprenants, de conception et annotation des corpus d'apprenants, de ressources de traitement automatique des langues naturelles appliquées à la recherche sur les apprenants, et d'applications didactiques de la recherche sur les corpus d'apprenants (Spina, 2017). Martelli (2008) a analysé les erreurs d'un corpus de productions écrites d'apprenants d'italien LE. Andorno (2005) a étudié l'enchâssement des particules additives et restrictives dans l'énoncé verbal en italien L2. Gallina (2010 ; 2013) a utilisé le corpus *LIPS*<sup>4</sup> pour l'étude de l'acquisition du lexique (richesse lexicale et champs lexicaux) en italien L2. Konecny *et al.* (2016) ont classé les phrasèmes dans un CA d'italien L2 (*LEKO*<sup>5</sup>) et proposé un modèle d'enseignement mixte des phrasèmes (Zanasi *et al.*, 2016). Bonsegna (2000) a étudié les faux amis italo-anglais dans la production écrite d'apprenants de niveau avancé d'italien LE.

La section suivante décrit une recherche menée par l'auteur (Casani, 2020) pendant le développement du projet *MERLIN*<sup>6</sup> (Université Technologique de Dresde, en partenariat avec l'Université Eberhard Karls Tübingen, l'*EURAC (European Academy)* de Bolzano/Bozen, et l'Université Charles IV Prague), un CA écrit trilingue (allemand, italien et tchèque) élicite par des apprenants de différentes langues maternelles.

## 2. Méthode

Nous avons analysé les textes écrits de 400 apprenants adultes d'italien LE (127 hommes et 273 femmes, âgés de 18 à 38 ans) du niveau A1 au niveau B2 du CECRL, de langue maternelle mixte avec une prévalence de germanophones (55%). Les autres langues maternelles sont le hongrois (17%), le polonais (13%), le français (10%), l'espagnol (1%), le portugais (1%) et d'autres (3%). Les textes constituent un sous-groupe de la section italienne du corpus MERLIN. Ils ont été élicites par des tests de certification du centre TELC<sup>7</sup> de Francfort, partenaire du gouvernement fédéral allemand et du projet MERLIN. Les tests ont été soumis aux procédures de contrôle de l'EALTA<sup>8</sup>, conformément aux normes internationales de qualité en matière de tests. Les tâches demandées aux apprenants d'italien LE pour chaque niveau sont les suivantes :

A1

- aider un ami à la recherche d'un emploi;
- déplacer un rendez-vous;

A2

- proposer à un ami de lui rendre visite ;
- contacter un ami après une longue période ;
- informer des amis d'un cours de langue auquel on participe ;

B1

- aider un ami à la recherche d'un emploi après l'obtention de son diplôme ;

<sup>3</sup> Pour une brève description des CA italiens et de leurs possibilités d'exploitation pédagogique, voir Casani (2018). Pour une liste complète mise à jour des CA existants et des sites des projets respectifs, voir <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>.

<sup>4</sup>*Lessico dell'Italiano Parlato da Stranieri* <http://www.parlitaliano.it/~parole/index.php/en/corpora/653-corpus-lips>.

<sup>5</sup>Lexemkombinationen und typisierte Rede im mehrsprachigen Kontext <http://www.leko-project.org/>.

<sup>6</sup>Mehrsprachige Plattform für die Europäischen Referenzniveaus <https://merlin-platform.eu/>.

<sup>7</sup>The European Language Certificates.

<sup>8</sup>European Association for Language Testing and Assessment.



- répondre à une lettre sur les plats typiques de son pays ;
- se renseigner sur un séjour d'étude ;
- répondre à une invitation de mariage ;

## B2

- postuler pour un stage dans une entreprise ;
- parler de ses expériences d'apprentissage des langues ;
- rejoindre un projet de solidarité ;
- rejoindre les soirées de cuisine internationales ;
- répondre à une lettre d'une personne qui a des problèmes avec les chats ;
- porter plainte contre un hôtel.

Pour compiler le corpus, les productions écrites ont été extraites des tests originaux et réévaluées par deux évaluateurs formés à l'utilisation de deux grilles d'analyse conformes au CECRL. La première grille est similaire au tableau 3 du CECRL (Council of Europe, 2001 : 29-30), adaptée aux besoins des évaluateurs (Alderson, 1991) et opérationnalisée ; la seconde est l'échelle d'évaluation globale du répertoire linguistique général (Council of Europe, 2001 : 110). Le degré de fiabilité des évaluations a été vérifié à l'aide de la théorie classique des tests et de l'analyse multi-facettes de Rasch. Pour chaque critère d'évaluation, les productions ont reçu un score directement orienté sur les niveaux du CECRL, ce qui a permis d'obtenir un profil de compétence pour chaque production. Il convient de noter que, à la suite des réévaluations, seuls deux textes évalués comme B2 par *TELC* ont obtenu un jugement pleinement conforme aux niveau B2, de sorte que le B2 traité ici ne correspond pas à un niveau d'utilisateur pleinement indépendant mais à un niveau de première autonomie (B2.1). Les textes ont été transcrits à l'aide de *XML-mind*<sup>9</sup>, un éditeur basé sur *XML*, par le personnel de *TELC*, qui a également veillé aux contrôles de fiabilité des transcriptions. Une fois la transcription terminée, les données ont été converties en *PAULA*<sup>10</sup>, un *XML* conçu comme format d'échange pour les annotations linguistiques. Pour les annotations manuelles, on a utilisé l'outil *Falko*<sup>11</sup>, un complément *Excel* développé par l'Université Humboldt de Berlin, pour les annotations des hypothèses cibles et l'outil *MMAX2*<sup>12</sup> pour les annotations multiniveaux des catégories linguistiques.

L'auteur a réalisé l'analyse décrite dans cette étude lors de la construction du corpus, à laquelle il a collaboré dans les années 2013 – 2014 à l'Université Technologique de Dresde. Les 400 textes, sélectionnés parmi ceux dans lesquels les annotations avaient été vérifiées, ont été importés dans *Annis*<sup>13</sup>, un navigateur web gratuit pour la recherche et la visualisation de corpus linguistiques multiniveaux complexes, et analysés selon les 15 paramètres morphosyntaxiques suivants, extraits du schéma d'annotation *MERLIN*<sup>14</sup>:

- ordre syntaxique: erreurs d'ordre des constituants en phrases principales (\*Mi molto dispiace au lieu de Mi dispiace molto) ou secondaires (\*perché secondo me questo è un positivo gesto au lieu de perché secondo me questo è un gesto positivo);
- négation: erreurs d'omission (\*lui viene mai più con noi in vacanza au lieu de lui non viene mai più con noi in vacanza), changement (\*no ho molto tempo au lieu de non ho molto tempo) et position (\*il mio luogo è non sviluppato

<sup>9</sup><https://www.xmlmind.com/>

<sup>10</sup>Potsdamer AUSTauschformat Linguistischer Annotationen [purl.org/net/paula](http://purl.org/net/paula)

<sup>11</sup>[purl.org/net/Falko](http://purl.org/net/Falko)

<sup>12</sup>[mmax2.net](http://mmax2.net)

<sup>13</sup><http://corpus-tools.org/annis/>

<sup>14</sup>Les grilles d'évaluation *MERLIN*, informations détaillées sur la préparation et l'annotation du corpus, ainsi qu'une vaste bibliographie sont disponibles sur le site web du projet.

molto au lieu de il mio luogo non è molto sviluppato) de particules de négation ;

- valence: erreurs d'omission (\*Hai superato con il massimo dei voti? au lieu de L'hai superato con il massimo dei voti?) ou addition (\*vorrei che mi vi rimborsiate au lieu de vorrei che mi rimborsiate) de constituants obligatoires ;
- accord: erreurs d'accord sujet-verbe (\*io vuole visitare Ville X<sup>15</sup> au lieu de io voglio visitare Città X) ;
- flexions inexistantes: utilisation de formes non existantes dans les paradigmes de flexion des noms (\**problemo* au lieu de *problema*), verbes (\**Sposarei in bianco* au lieu de *Sposerei in bianco*) ou adjectifs (\**una breva lettera* au lieu de *una breve lettera*) ;
- morphologie nominale: erreurs de genre (\**i mani* au lieu de *le mani*), nombre (\**Alle fine* au lieu de *Alla fine*) et cas (\**persone come tu* au lieu de *persone come te*) dans les parties nominales de la phrase ;
- morphologie verbale: erreurs de mode (\*So, che fra poco, avresti un proprio appartamento au lieu de So che fra poco avrai un proprio appartamento), temps (\*Quando siamo ritornati, è possibile incontrarti? au lieu de Quando saremo ritornati, è possibile incontrarti?) et aspect (\*Il mio esame era abbastanza bene au lieu de Il mio esame è stato abbastanza buono) des verbes ;
- verbe principal: addition (\*Andiamo tornare a casa au lieu de Andiamo a casa) et omission (\*Un anno fa in Germania per una settimana au lieu de Un anno fa sono stato in Germania per una settimana) du verbe principal ;
- prédicats analytiques: erreurs dans la formation de prédicats complexes impliquant omission (\**Michele e Luca andati in città* au lieu de *Michele e Luca sono andati in città*), addition (\**Mentre ho leggevo il giornale* au lieu de *Mentre leggevo il giornale*), modification/substitution (\**siamo potuti giocare* au lieu de *abbiamo potuto giocare*) de verbes auxiliaires, modaux, causatifs et de copulas ;
- conjonctions: erreurs d'omission (\*Speriamo tu stai bene au lieu de Speriamo che tu stia bene), addition (\*da due anni che sto imparando inglese au lieu de da due anni sto imparando l'inglese), substitution (\*Sarà meglio quando uno di voi mi chiama au lieu de Sarà meglio se uno di voi mi chiama) et position (\*e poi vuole fare la maturità anche au lieu de e poi vuole fare anche la maturità) des conjonctions ;
- pronoms clitiques et réfléchis: erreurs d'omission (\*Vi scrivo per lamentare au lieu de Vi scrivo per lamentarmi), addition (\*Come lo sai, avevo qualche problema au lieu de Come sai, avevo qualche problema), malformation (\*se trova entro il Lago Maggiore au lieu de si trova entro il Lago Maggiore) et position (\*puoi la chiamare au lieu de puoi chiamarla) dans les pronoms clitiques et réfléchis ;
- parties du discours: erreurs de substitution de parties du discours (\**la città è molto bene* au lieu de *la città è molto bella*) ;
- prépositions : erreurs d'omission (\*giocare tennis au lieu de giocare a tennis),

<sup>15</sup> Les noms de personnes et de villes ont été anonymisés.

addition (\*Vorremmo di un parziale rimborso au lieu de Vorremmo un parziale rimborso) et substitution (\*Le scrivo al riferimento au lieu de Le scrivo in riferimento) des prépositions ;

- articles: erreurs d'omission (\*Lavoro anche in ristorante dei miei genitori au lieu de Lavoro anche nel ristorante dei miei genitori), addition (\*In riferimento al questo annuncio au lieu de In riferimento a questo annuncio) et substitution (\*i appartamenti au lieu de gli appartamenti) des articles.

Les analyses inférentielles de l'association entre niveau et type d'erreur et des différences de proportions d'erreurs entre niveaux de compétence ont été effectuées à l'aide de *IBM-SPSS24*.

### 3. Résultats

Le tableau 1 montre les proportions d'erreurs (intra-niveau) par niveau de compétence et catégorie grammaticale. L'association entre niveaux et catégories d'erreur est significative ( $\chi^2(39) = 116.017, p < 0.001$ ). Les résultats de l'analyse *post hoc* (*Z*) sont représentés au moyen de couleurs différentes. Dans la même catégorie linguistique, les cases de couleur sombre indiquent des proportions d'erreur statistiquement plus élevées que celles de couleur claire ; les cases non peintes indiquent des proportions qui ne diffèrent pas de manière statistiquement significative ( $p \leq 0.05$ ) des autres dans la même colonne.

Les tableaux 2, 3 et 4 montrent les proportions d'erreur (intra-niveau) groupées respectivement par morphologie nominale, morphologie verbale et construction syntaxique. Ces tableaux incluent les erreurs superficielles, c'est-à-dire les modifications des structures au niveau de la surface (addition, omission, changement, position). L'association entre niveaux et catégories d'erreur est statistiquement significative (*Fisher* = 275.411,  $p < 0.001$ ; *V* = 0.237,  $p < 0.001$ ). Les résultats de l'analyse *post hoc* (*Z*) ( $p \leq 0.05$ ) sont représentés par les différentes couleurs (voir ci-dessus) dans les tableaux.

### 4. Courte discussion<sup>16</sup>

La plupart des catégories linguistiques ont tendance à la distribution normale des erreurs, avec un pic entre le niveau de survie et le niveau seuil. Ce cadre est compatible avec une introduction progressive des structures, qui sont absentes ou non-analysées dans le niveau A1, sont expérimentées dans les niveaux de survie et seuil, puis ont tendance à se consolider au fur et à mesure que l'interlangue évolue vers l'indépendance. Les exceptions sont certaines catégories qui présentent un maximum d'erreurs dans le niveau A1. Celles-ci diminuent dans l'A2 en proportion statistiquement significative pour la flexion inexistante des verbes, pour l'omission de parties des prédicats analytiques, pour les omissions de pronoms réfléchis et pour la substitution de parties du discours. La distribution des erreurs dans ces catégories tend alors à se normaliser aux niveaux suivants, sauf dans les prédicats analytiques, qui voient une nouvelle diminution numérique au B1 et une augmentation significative au B2. Dans le cas de telles tendances irrégulières, l'analyse de sous-catégories supplémentaires au niveau superficiel peut suggérer un ajustement possible des étiquettes d'annotation. Dans les prédicats analytiques, par exemple, seulement les omissions (principalement des copules et des verbes modaux) diminuent de manière significative dans le niveau B1, atteignant presque zéro, tandis que les malformations restent statistiquement constantes, ainsi que les ajouts, qui restent proches de zéro. Les erreurs de morphologie nominale (14,5%) sont présentes dans une proportion prépondérante par rapport à celles de morphologie verbale (8,8%) et restent statistiquement inchangées à tous les niveaux, contrairement aux erreurs verbales, qui

<sup>16</sup> Pour une discussion détaillée des données accompagnée d'exemples extraits du corpus, voir Casani (2020).

augmentent de manière statistiquement significative entre le niveau de découverte et le niveau de survie, parallèle à l'élargissement du répertoire des temps et des modes verbaux. Les plus grandes proportions d'erreurs se produisent dans les prépositions (20,7%) et les articles (15%), mais si les prépositions restent problématiques à tous les niveaux de compétence (avec une augmentation significative du nombre de substitutions au niveau A2), les articles sont soumis à une amélioration significative des omissions lors du passage au niveau seuil, ce qui conduit ensuite à une certaine tendance à la surutilisation au niveau B2. Le passage de phrases courtes juxtaposées à des expressions plus complexes entraîne une augmentation significative de l'omission des conjonctions au niveau A2, tandis que l'expérimentation de l'hypotaxe au niveau B2 entraîne une augmentation significative du nombre de substitutions de conjonctions. Il n'y a pas de variations significatives d'un niveau à l'autre en ce qui concerne les erreurs de distribution syntaxique (3,9%), de négation (0,8%), de valence (5%), d'accord sujet-verbe (2,9%) et d'ajout/omission du verbe principal (1,6%).

	Ordre synt.	Neg.	Val.	Acc. S/V	Flex. inex.	Morph. nomin.	Morph. verb.	Verbe princ.	Predic. analytiques	Conjonct.	Clit.+ refl.	Parties du disc.	Prep.	Art.
B2.1	3,2	0,6	4,1	3,8	3,5	16,4	9,8	1,9	6	6,9	6	3,8	17,7	16,4
B1	4,4	0,6	4,8	3,3	4,2	14,1	10,6	1,2	2,3	4,8	9,6	5	23,9	11,2
A2	4,1	1	4,1	3,3	3,3	15,6	11,1	0,8	3,3	6	3,9	3,9	21,6	17,9
A1	3,9	1	6,9	1,3	8	12,1	3,1	2,6	7,7	2,1	9,8	8,5	18,3	14,9

Tableau 1 - Erreurs par niveau et catégorie grammaticale (% intra-niveau)

	Flex. Inexist.	Aspect	Pred. analyt. (add)	Pred. analyt. (change)	Pred. analyt. (omiss)	Verbe princ. (add)	Verbe princ. (omiss)	Mode	Temps	Accord
B2.1	0	2,5	0	4,4	1,6	0,3	1,6	4,1	3,2	3,8
B1	1,9	1	0,4	1,7	0,2	0	1,3	6,5	3,1	3,3
A2	1,2	2,4	0	1,6	1,6	0	0,8	5,5	3,1	3,3
A1	4,4	0	0,3	4,1	3,3	0,3	2,3	2,6	0,5	1,3

Tableau 2 - Erreurs de morphologie verbale par niveau de compétence et catégorie linguistique (% intra-niveau)

	Flex. inex. (nom)	Flex. inex. (adj)	Articl. (add)	Articl. (cha)	Articl. (omiss)	Clit. (add)	Clit. (chan)	Clit. (omiss)	Clit. (pos)	Refl. (add)	Refl. (chan)	Refl. (omiss)	Refl. (pos)
B2.1	3,5	0	7,6	1,9	6,9	1,3	0,6	3,2	0	0,6	0,3	0	0
B1	1,9	0,4	5,0	2,7	3,6	1,7	1,5	2,7	0,8	0	0,4	1,7	0,2
A2	1,8	0,2	4,9	1,0	11,4	1,6	0,6	1,0	0,2	0	0,2	0,2	0
A1	2,3	1,3	2,8	0,5	11,6	0,5	1,5	2,3	1,3	0	0,8	3,3	0

Tableau 3 - Erreurs de morphologie nominale par niveau de compétence et catégorie linguistique (% intra-niveau)

	Ordre syntax. (phrase princ.)	Ordre syntax. (phrase subordonnée)	Clitiques (pos.)	Réfléchis (pos)	Conjonctions (pos)	Negations (pos)
B2.1	1,9	1,3	0	0	0,6	0,3
B1	3,1	1,3	0,8	0,2	0,8	0
A2	3,3	0,8	0,2	0	1,4	0
A1	3,1	0,8	1,3	0	0,5	0,3

**Tableau 4 – Erreurs de structure syntaxique par niveau de compétence et catégorie linguistique (% intra-niveau)**

Les tableaux 1, 2, 3 et 4 peuvent être utilisés comme support pour l'évaluation de la morphosyntaxe des productions écrites d'apprenants d'italien LE ou pour l'étude des interlangues. L'utilisation peut être faite à un niveau plus général, en considérant uniquement le changement de couleur dans chaque catégorie linguistique. Dans la même colonne, la transition du gris foncé au gris clair indique une amélioration significative de la catégorie linguistique respective. Inversement, le passage du gris clair au gris sombre indique une détérioration significative de la catégorie linguistique concernée. La couleur blanche n'indique aucun changement significatif. À un niveau plus fin, les pourcentages peuvent donner une idée de la proportion d'erreurs pouvant être attendue à un certain niveau pour chaque catégorie linguistique. Les grilles peuvent être composées en fonction des besoins, en regroupant les catégories utiles, et mises en œuvre via l'analyse d'erreurs des CA. Les données statistiques peuvent atteindre un niveau de généralisation plus élevé grâce à la composition de grilles verbales combinant des descriptions quantitatives et qualitatives des erreurs, structurées selon les critères de formulation positive, de concision et de transparence du CECRL. En annexe on propose des descripteurs pour la compétence morphosyntaxique générale (annexe 1), pour la morphologie nominale (annexe 2) et verbale (annexe 3), ainsi que pour la construction syntaxique (annexe 4), résultats des analyses présentés dans cette étude. Ils peuvent être accompagnés d'exemples authentiques tirés des CA de référence à la fois pour les structures individuelles et pour les textes intégraux<sup>1</sup>. Ces descripteurs constituent une réponse possible aux recommandations du Conseil de l'Europe qui, depuis la publication du CECRL, a encouragé le développement d'outils d'aide au Cadre pour la description empirique des niveaux de compétence dans les différentes langues communautaires (voir § 1).

## 5. Conclusions

Au cours des trois dernières décennies, les corpus d'apprenants ont acquis un rôle de premier plan dans le domaine de l'acquisition de la L2/LE. À l'heure actuelle, il existe cependant encore peu d'études concernant l'application à la pratique réelle des résultats des CA italiens et l'italien est encore sous-représenté en termes de recherche appliquée aux CA.

Cette contribution illustre une application possible de la méthodologie d'analyse d'erreurs à un CA à travers une étude visant à la validation empirique des niveaux du CECRL pour la compétence morphosyntaxique en italien LE. Les résultats ont été synthétisés dans des grilles construites sur le modèle des descripteurs du CECRL. Ces grilles sont présentées en annexe comme une aide pour l'évaluation de la morphosyntaxe dans les productions écrites des apprenants d'italiens LE (Casani, 2020).

<sup>1</sup> Pour des exemples de différentes structures, voir Casani (2020). Pour consulter et télécharger les exemples ainsi que les textes complets, voir <https://merlin-platform.eu/>.

### **Références bibliographiques**

- ALDERSON, J. Charles (1991). Bands and scores, in *Language testing in the 1990s*, Alderson, J. Charles & North, Brian (Eds.). London: Modern English Publications and the British Council, 71-86.
- ANDORNO, Cecilia (2005). Additive and restrictive particles in Italian as a Second Language. Embedding in the verbal utterance structure, in *The structure of learner varieties*, Hendriks, Henriëtte (Ed.). Berlin & New York: Mouton de Gruyter, 405-444.
- BONSEGNA, Chiara (2000). Italian-English false friends in the written production of advanced Italian EFL learners: a corpus-based analysis. Unpublished Thesis. Torino: Università di Torino.
- CARLSEN, Cecilie H. (2013). *Norsk Profil. Det europeiske rammeverket spesifisert for norsk. Et første steg*. Oslo: Novus forlag.
- CASANI, Emanuele (2018). Corpus linguistics e didattica dell'italiano. I learner corpora nella classe di lingue, in *Nella classe di italiano come lingua seconda/straniera*, D'Angelo, Maria Carmela & Diadori, Pierangela (Eds). Firenze: Franco Cesati Editore, 157-167.
- CASANI, Emanuele (2020). Valutare la competenza morfosintattica in italiano L2. Una validazione corpus-based dei livelli del QCER, in *Valutazione e misurazione delle produzioni orali e scritte in italiano lingua seconda*, Nuzzo, Elena, Santoro, Elisabetta & Vedder, Ineke (Eds). Firenze: Franco Cesati Editore.
- COUNCIL OF EUROPE (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- COUNCIL OF EUROPE (2005). *Reference level Description for National and regional Languages. Guide for the production of RLD*. Strasbourg: Council of Europe Language Policy Division, DG IV.
- GALLINA, Francesca (2010). The LIPS Corpus (Lexicon of Spoken Italian by Foreigners) and the acquisition of vocabulary by learners of Italian as L2, in *Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching, vol. 4: Papers from LAEL PG 2009*, Bota, Grace, Hargreaves, Helen, Chia-Chun, Lai & Rong, Rong (Eds). Lancaster, Lancaster University, 30-50.
- GALLINA, Francesca (2013). The Lexicon of Spoken Italian by Foreigners: A study on the acquisition of vocabulary by L2 Italian learners between measures of lexical richness and lexical fields, in *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead. Proceedings of the First Learner Corpus Research Conference (LCR 2011)*, Granger, Sylviane, Gilquin, Gaëtanelle & Meunier, Fanny (Eds). Louvain-la-Neuve: Presses Universitaires de Louvain, 179-195.
- GRANGER, Sylviane (2002). A bird's-eye view of learner corpus research, in *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, Granger, Sylviane, Hung, Joseph & Petch-Tyson, Stephanie (Eds). Amsterdam/Philadelphia: John Benjamins, 3-33.
- GRANGER, Sylviane (2003). Error-tagged learner corpora and CALL: a promising synergy. Special issue on error analysis and error correction in computer-assisted language learning. *CALICO Journal*, 20 (3): 465-480.
- HANA, Jirka, ROSEN, Alexandr, SVATAVA, Skodov' & STINDLOVÁ, Barbora (2010). Error-tagged Learner Corpus of Czech, in *Proceedings of the Fourth Linguistic Annotation Workshop, LAW 2010*, 11-19.
- HULSTIJN, Jan H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91: 663-667.
- KONECNY, Christine, ABEL, Andrea, AUTELLI, Erica & ZANASI, Lorenzo (2016). Identification and Classification of Phrasemes in an L2 Learner Corpus of Italian, in *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives -*

*Fraseologia computacional y basada en corpus: perspectivas monolingües y multilingües.* Geneva: Tradulex, 533–542.

LEECH, Geoffrey (1998). Learner corpora: what they are and what can be done with them, in *Learner English on Computer*, Granger, Sylviane (Ed). London/New York: Longman, xiv-xx.

MARTELLI, Aurelia (2008). Error analysis and learner corpora: a study of errors in the written production by English students of Italian, in *Investigating English with corpora. Studies in honour of Maria Teresa Prat*, Martelli, Aurelia & Pulcini, Virginia (Eds). Milano: Polimetrica Publisher, 365–378.

SPINA, Stefania (2017). Learner Corpus Research and the acquisition of Italian as a second language: the case of the Longitudinal Corpus of Chinese Learners of Italian (LoCCLI), in *4th Learner Corpus Research Conference. LCR 2017. Book of Abstracts*, Abel, Andrea (Ed). Bolzano/Bozen: Eurac Research, 18-19.

SPINELLI, BARBARA & PARIZZI, Francesca (2010). *Profilo della lingua italiana*. Firenze: La Nuova Italia.

TAN, Melinda (2005). Authentic language or language errors? Lessons from a learner corpus, *ELT Journal*, 59 (2), 26–34.

TONO, Yukio (2003). Learner corpora: Design, development and applications. In *Proceedings of the 2003 Corpus Linguistics Conference*, Archer, Dawn Elizabeth, Rayson, Paul, Wilson A. & McEnery, Tony (Eds). UCREL: Lancaster University, 800–809.

ZANASI, Lorenzo, KONECNY, Christine, AUTELLI, Erica & ABEL, Andrea (2016). L'insegnamento dei fraseologismi nell'italiano come lingua seconda: proposta di un modello didattico misto, in *Lingua al plurale: la formazione degli insegnanti*, De Marco, Anna (Ed). Perugia: Guerra, 171-180.

## Annexes

Les descripteurs structurés à la suite des présentes analyses (Casani, 2020) sont mentionnés ci-dessous.

<b>Morphosyntaxe générale</b>	
<b>B2.1</b>	Il fait preuve d'une bonne maîtrise grammaticale. Les prédicats sont généralement complets, mais ils continuent à faire des erreurs, évidentes dans les constructions analytiques, en particulier avec les participes et les modaux. Ces constructions sont cependant généralement complètes. Les erreurs de morphologie nominale sont encore fréquentes, en particulier dans le genre, mais il ne commet pas d'erreurs pouvant provoquer des malentendus.
<b>B1</b>	Il utilise de manière raisonnablement correcte des formules de routine et des structures fréquemment utilisées, liées aux situations les plus prévisibles. Il formule des prédicats généralement complets dans lesquels des omissions des actants continuent à se produire, généralement dans les références pronominales, mais l'ajout de compléments superflus est rare. Malgré quelques erreurs, ce qu'il communique est généralement clair.
<b>A2</b>	Il utilise correctement certaines structures simples, dans lesquelles il exprime les prédicats de manière assez complète, omettant parfois les actants et commettant des erreurs systématiques de base. Par exemple, il confond les temps et les modes verbaux et il oublie de signaler les accords, mais les formes inexistantes diminuent et les parties du discours se distinguent généralement.
<b>A1</b>	Il possède une maîtrise limitée de structures grammaticales simples et de modèles syntaxiques simples dans un répertoire stocké. Il construit des prédicats partiels en omettant les actants et en ajoutant ceux qui sont superflus. Il utilise souvent des noms et des verbes de manière indistincte et expérimente leur flexion, créant souvent des formes personnelles.

Annexe 1 – Exemples de descripteurs de la compétence morphosyntaxique générale



<b>Morphosyntaxe nominale</b>	
<b>B2.1</b>	Bien que des difficultés persistent avec le genre et le nombre, il utilise des constructions généralement claires et plus complexes, en utilisant des références pronominales dans lesquelles il fait parfois des erreurs
<b>B1</b>	Ce qu'il communique est généralement clair. Il fait un usage intensif des articles, mais il a toujours des difficultés avec l'accord de genre et de nombre.
<b>A2</b>	Les formes inventées diminuent, bien que des erreurs de genre et de nombre persistent.
<b>A1</b>	Il fait un usage personnel de la flexion, quels que soient les accords.

**Annexe 2 - Exemples de descripteurs de la morphologie nominale**

<b>Morphosyntaxe verbale</b>	
<b>B2.1</b>	Il utilise toujours correctement des formes verbales existantes, même s'il fait encore des erreurs, par exemple en remplaçant des auxiliaires, ce qui ne compromet pas la compréhension du message. Des erreurs de temps et d'aspect persistent également, souvent dues à l'influence de la langue maternelle.
<b>B1</b>	Il fait un usage systématique de l'aspect verbal et exprime la modalité d'une manière assez claire, mais il continue de commettre des erreurs dans les temps, en particulier dans l'accord de formes perfectives.
<b>A2</b>	Il utilise presque toujours des formes verbales existantes, bien qu'il fasse encore des erreurs d'accord, des omissions d'auxiliaires et de modaux. Bien qu'avec des erreurs, il commence à exprimer des différences d'aspect et des modes différents de l'indicatif.
<b>A1</b>	Il utilise souvent des formes verbales inexistantes.

**Annexe 3 - Exemples de descripteurs de la morphologie verbale**

<b>Construction syntaxique</b>	
<b>B2.1</b>	Il utilise des constructions assez correctes, aussi hypothétiques, dans lesquelles on trouve parfois des adverbes et des connecteurs dans une position non standard.
<b>B1</b>	Il utilise des constructions syntaxiques plus articulées, introduisant des phrases secondaires (objectives, temporelles et certaines relatives) dans lesquelles cependant l'ordre des constituants ne correspond parfois pas à la norme.
<b>A2</b>	Il dispose d'un petit répertoire linguistique dans lequel des constructions atypiques se produisent en raison d'échanges nom/adjectif, verbe/adverbe et position de la conjonction <i>anche</i> .
<b>A1</b>	Il utilise des phrases courtes, dans lesquelles les possibilités de constructions incorrectes sont plutôt limitées. Cependant, des échanges sujet/verbe et des positions incorrectes des adverbes ( <i>molto, sempre, insieme</i> ) se produisent.

**Annexe 4 - Exemples de descripteurs de la construction syntaxique**

# APPROCHER LES TEXTES PAR LES NOTIONS DE « GENRES » ET DE « TYPES DE DISCOURS » : CHOIX DIDACTIQUES POUR LA CONCRETISATION AU SEIN D'UN ENVIRONNEMENT NUMERIQUE

Triscia BIAGIOTTI<sup>a</sup>, Giada QUARANTA<sup>b</sup>

[triscia.biagiotti@univ-grenoble-alpes.fr](mailto:triscia.biagiotti@univ-grenoble-alpes.fr), [giadaquaranta@gmail.com](mailto:giadaquaranta@gmail.com)

<sup>a,b</sup> *Projet IDEFI Innovalangues, Maison des Langues et des Cultures, Université Grenoble Alpes, Grenoble, France*

## 1. Introduction

La présente contribution se donne pour but de poser les jalons d'une approche didactique par « genres de texte » et « types de discours » pour l'apprentissage des langues-cultures (notamment l'italien et l'anglais). Cette approche a été configurée dans le cadre du projet IDEFI Innovalangues (ANR-11-IDEFI-0024) et notamment dans la conception des Parcours d'Apprentissage en Autonomie (PAA), réalisations didactiques numériques visant le renforcement de la compréhension de l'oral en langues en contexte universitaire, construites autour de séquences vidéo à forte valeur culturelle, intégrées dans un environnement numérique *ad hoc*<sup>1</sup>. Les PAA se proposent de renforcer la compréhension de l'oral à travers une approche didactique de la langue qui part de la globalité du texte et s'intéresse ensuite à ses constituants à différents niveaux, en adoptant un grain d'analyse de plus en plus fin.

Nous commencerons par illustrer le cadre théorique dans lequel s'inscrit notre conception des notions de « genre de texte » et de « types de discours ». Nous discuterons également les recherches en didactique qui ont repensé les notions en question pour les réinvestir notamment dans le champ de l'enseignement-apprentissage du français langue première et seconde (Schneuwly & Dolz, 1997 ; Chartrand, 2015, 2016 ; Bemporad & Pahud, 2016 ; Miranda & Coutinho, 2015). Nous terminerons cette première partie par l'exposé des critères que nous avons adoptés pour concevoir les PAA.

La deuxième partie présentera la structure globale des parcours ainsi qu'une unité prototypique conçue pour les apprenants d'italien de niveau B2. L'illustration de cette unité nous permettra, en particulier, de montrer les potentialités des outils techno-pédagogiques que nous avons conçus afin de proposer un travail de compréhension profonde du texte oral et qui aide les apprenants à se focaliser sur le processus plus que sur le résultat de la compréhension. Enfin, nous terminerons en citant quelques résultats issus d'une première expérimentation menée à titre exploratoire qui nous a permis d'obtenir un premier retour des étudiants et d'élargir l'éventail de nos pistes de réflexion.

## 2. Cadre théorique

Les notions de « genres de texte » et de « type de discours » sont présentes dans le CECRL (pp 75-80)<sup>2</sup>, mais il a fallu les repenser pour les rendre opératoires en vue de nos objectifs

<sup>1</sup> Cette recherche a été financée dans le cadre du projet IDEFI Innovalangues (ANR-11-IDEFI-0024), sous la direction scientifique de Monica Masperi (LIDILEM, Université Grenoble Alpes). Sandra Canelas-Trevisi (GRAFE, Université de Genève), a joué un rôle déterminant dans la définition des choix théoriques et méthodologiques et dans la réalisation des PAA (Canelas-Trevisi : 2009, 2015 et 2019).

<sup>2</sup> Le Cadre présente une liste non exhaustive de « genres et types de textes » (CECRL, 2001 : 75-81) oraux et écrits qui sont susceptibles d'être associés à des activités d'apprentissage. Cependant, aucune indication précise n'est fournie concernant une approche didactique fondée sur les notions de genres de textes et types de discours.

spécifiques. Nous avons été amenées en effet à établir des critères susceptibles d'être utilisés aussi bien dans la sélection des ressources que dans la construction didactique des parcours à soumettre aux étudiants.

La discussion critique d'une notion aussi complexe que le genre dépasse largement le cadre restreint de cet article et les objectifs qu'il poursuit. Nous présentons donc directement le cadre théorique dans lequel nous nous inscrivons. Nos objectifs nous ont orientés vers des recherches que nous pouvons distinguer synthétiquement en deux grands ensembles : celles qui sont ancrées dans l'étude du fonctionnement langagier, et de l'activité humaine (Bronckart, 1997, 2004 ; Rastier, 2001) et celles qui émanent plutôt de l'étude et de l'analyse des textes (Adam, 1992, 2011, 2015) et en identifient les caractéristiques comme des traces d'influences sociales et culturelles.

Nous avons exploré en parallèle avec les recherches théoriques les recherches didactiques qui s'y réfèrent et qui en ont reconfiguré les principes en prenant en compte des contextes d'enseignement et d'apprentissage. C'est dans les travaux de Suzanne Chartrand (2008, 2013 et 2015) notamment, que nous avons puisé le principe de la reconfiguration des notions de genre et de type, développés ci-dessous.

### 2.1. Les concepts de « genres de texte » et de « types de discours »

La conception de parcours basés sur une approche par « genres de texte » et par « types de discours » place la notion de « texte » au centre de nos réflexions. Dans le droit fil des travaux d'Adam (1987, 1992, 1997, 2005, 2011, 2015) et de Bronckart (1996, 1997, 2004, 2008), nous avons choisi d'appréhender le texte en tant que résultat observable de l'activité langagière et de le mettre en regard d'une langue-culture donnée. Selon cette approche, l'activité langagière se situe à l'intérieur des pratiques sociales en contexte (Bronckart, 1997 et 2004).

En partant du principe qu'une *science du langage* doit considérer avant tout « les productions verbales effectives, dans leur contexte, c'est-à-dire les textes », Bronckart considère les pratiques sociales dans lesquelles sont conçus, produits et reçus les textes comme « éminemment variables » (Bronckart, 2008b : 10). Cette variabilité des pratiques sociales fait que les textes « se différencient eux-mêmes en de multiples « formes standard » ou en genres textuels » (ibid.).

Pour cet auteur, tout texte est produit, reçu et perçu dans un contexte qui comporte une dimension matérielle et une dimension socio-subjective. Tout texte empirique est ainsi lié à une situation matérielle de production, au cours de laquelle l'énonciateur, individu physique, s'adresse à un ou plusieurs autres individus, les destinataires, en produisant un texte dans un lieu et un temps déterminé : le lieu et le temps de production. En plus de la situation matérielle, le texte est également lié à une situation d'interaction sociale définie à travers les paramètres suivants :

- le lieu social : un espace social défini par des attentes et des normes plus ou moins stabilisées ;
- les individus impliqués dans l'interaction : l'énonciateur et le destinataire, rôles sociaux qui se définissent dans une pratique sociale à un moment donné de l'histoire ;
- les projections de l'effet que le texte aura sur le destinataire. Ce domaine complexe de calculs, plus ou moins maîtrisés, ancré dans le monde social et subjectif de l'énonciateur et du destinataire, est désigné par le terme de « but du texte ».

La dimension socio-subjective du contexte comporte ainsi les connaissances encyclopédiques et sociales que chaque membre d'une société donnée construit au cours de son existence. Tout

locuteur d'une langue donnée dispose d'une certaine connaissance des genres et le choix du genre qu'il considère comme le mieux adapté au contexte social de production, au destinataire et au but visé, représente une manifestation de ces connaissances.

Selon Adam, « Les genres, organisés en système de genres, sont des patrons sociocommunicatifs et sociohistoriques que les groupes sociaux se donnent pour organiser les formes de la langue en discours. » (Adam, 2011 : 33). En définitive, tout texte donne à voir un effet de généricité et il ne peut y avoir de textes sans genre (ibid.).

Les genres sont perçus par les locuteurs comme des ensembles de textes identifiés à partir de caractéristiques hétérogènes – but communicatif, pratique sociale et lieu social, caractère plus ou moins codifié – et ils sont désignés à travers des termes tout aussi hétérogènes : interview, conférence, reportage, nouvelle, plaidoyer... Ainsi, la notion de genre désigne-t-elle « une manière de faire les textes » qui s'est stabilisée au cours de l'histoire, mais qui est en transformation permanente, sous l'action des forces qui interagissent dans l'espace social.

D'après Adam (2011) « Tout texte est la trace langagière d'une interaction sociale, la matérialisation sémiotique d'une action sociohistorique de parole » (2011 : 33). Le texte est donc ancré dans un contexte sociohistorique précis. Par ailleurs, en se référant au contexte, l'auteur observe que « nous n'avons pas accès au contexte comme donnée extralinguistique objective, mais seulement à des (re)constructions par des sujets parlants et/ou par des analystes [...]. Les informations du contexte sont traitées sur la base des connaissances encyclopédiques des sujets, de leur préconstruits culturels. » (2006 : 4). Chaque individu produit et reçoit ainsi un texte sur la base des connaissances qu'il a et qu'il partage plus ou moins largement avec les autres membres de la société dans laquelle il vit et interagit.

Plusieurs auteurs ont travaillé sur les différentes typologies de textes et ils ont abouti à des propositions de catégorisation. Pour notre travail, nous faisons référence principalement à la catégorisation adoptée par Adam (1997) et aux travaux de Bronckart (2008).

Pour Adam (2005), le texte se compose de séquences ; chaque séquence comprend des caractéristiques formelles propres et elle est porteuse de valeurs discursives. Elle est le produit d'opérations cognitives qui organisent le texte dans son ensemble, en fonction du but visé (Monte, 2014). Adam identifie sept séquences qu'il nomme séquentialités :

1. narrative
2. descriptive
3. argumentative
4. expositive-explicative
5. injonctive-instructionnelle
6. conversationnelle
7. poétique-autotélique

Toutefois, tout texte étant caractérisé par une certaine hétérogénéité, Adam (1987) préfère parler de « dominante ». Tout comme Adam, Bronckart considère que les textes sont inscrits dans des genres, qu'ils sont pour l'essentiel constitués de segments différents qui présentent des configurations spécifiques d'unités linguistiques (2008). C'est sur le statut de ces segments que leurs positions diffèrent. Bronckart qualifie les segments de textes de « types de discours » : discours interactif, discours théorique, récit interactif, narration. Sa notion de type de discours se fonde sur les modalités d'organisation de l'activité psycho-langagière en termes d'opération psychologiques ayant trait au traitement du contexte et du contenu. Alors que dans l'usage courant du terme de discours, les sphères d'activités dans les pratiques sociales déterminent les choix discursifs, pour Bronckart, le rapport aux sphères d'activité est englobant et détermine le choix du genre, alors que le mode de structuration de l'action langagière est englobé et détermine le choix des segments de texte (Monte et Philippe, 2014).

### 3. Reconfiguration des concepts dans le champ de la recherche en didactique

Tout texte, à l'oral comme à l'écrit<sup>3</sup>, peut être associé à un genre, certes de manière plus ou moins tranchée, à partir de la pratique sociale dont il émane. Nous partons du principe que dans toutes les langues-cultures les locuteurs peuvent associer des textes à des genres. Ainsi, cette appartenance d'un texte à un genre est-elle pour ainsi dire un concept « universel ». Reconfigurer le concept de « genre de texte » afin de l'employer dans notre projet d'enseignement-apprentissage nous a donc permis de repenser ce qu'il y a d'universel dans le fonctionnement du langage, ce qui « traverse » des langues-cultures différentes et même typologiquement distantes.

En ce qui concerne la reconfiguration sur le plan didactique de nos fondements théoriques de référence, nous avons retenus les contributions présentes en milieu francophone. Selon Schneuwly et Dolz (1997 : 29) « C'est au travers des genres que les pratiques langagières s'incarnent dans les activités des apprenants » ; le concept de « genre » devient central dans l'enseignement du français langue maternelle.

La conception et la construction des Parcours d'Apprentissage en Autonomie a notamment pris appui sur les travaux de Suzanne Chartrand, didacticienne québécoise (2013, 2015), qui définit le *genre* comme « [...] un ensemble de productions langagières orales ou écrites qui, dans une culture donnée, possèdent des caractéristiques communes d'ordres communicationnel, textuel, sémantique, grammatical, graphique ou visuel et/ou d'oralité, souples mais relativement stables dans le temps » (Chartrand & alii, 2015 : 3).

Le genre apparaît donc comme une caractéristique intrinsèque de toutes les productions langagières auxquelles sont exposés les élèves, à l'école comme dans tous les autres contextes sociaux (Chartrand, 2016 : 63). Cette vision de Chartrand renvoie à la conception de Schneuwly (2000 : 32, 35) d'après qui le genre est un « méga-outil sémiotique » car il comprend différents outils sémiotiques, certains spécifiques aux genres et d'autres plus techniques (à savoir, l'orthographe et la syntaxe) (ibid.).

### 4. La construction didactique des PAA

La conception et la construction des PAA selon l'approche retenue s'est fondée sur un ensemble de critères organisés par deux grands principes : le premier a orienté la sélection des ressources audiovisuelles à soumettre aux étudiants, le deuxième a permis de définir des procédures didactiques susceptibles de faciliter l'accès aux contenus proposés. Ces procédures ont été configurées pendant la conception de nos activités et ensuite elles ont été reconfigurées pour être intégrées dans un Environnement Numérique Personnalisé d'Apprentissage (ENPA, Caroline Connect<sup>4</sup>).

#### 4.1. Les objectifs des parcours

Les PAA sont destinés à un public hétérogène d'étudiants<sup>5</sup> et proposent un renforcement de la compréhension de l'oral en langue-culture cible (dans notre cas, de l'italien) à travers une modalité d'apprentissage principalement à distance.

<sup>3</sup> « On appelle « texte » toute séquence discursive orale ou écrite que les usagers/apprenants reçoivent, produisent ou échangent. En conséquence, il ne saurait y avoir acte de communication langagière sans texte » (Conseil de l'Europe, 2001 : 75).

<sup>4</sup> La plateforme Caroline Connect est en évolution constante. La version qui héberge actuellement les PAA est la V10, disponible en ligne à l'adresse suivante : <https://claco-innovalangues.univ-lyon1.fr/>

<sup>5</sup> Des étudiants inscrits à une formation universitaire (en licence, en master ou en doctorat), non spécialistes (secteur LANSAD) et spécialistes (LLECR, LEA) qui souhaitent améliorer et approfondir leurs connaissances et compétences en langue-culture.

A partir du principe que tout locuteur possède une représentation plus ou moins étendue des genres disponibles dans sa langue, notre but est d'activer chez l'étudiant sa capacité à s'appuyer sur la connaissance des genres qu'il possède dans sa/ses langue(s)-culture(s) de référence ainsi que sur celle qu'il est en train d'échafauder dans la langue-culture cible. Il s'agit d'amener les étudiants à s'interroger sur le genre de vidéos qui leur sont proposées dans chaque parcours afin qu'ils puissent *consciemment* adopter la posture qu'ils adoptent habituellement face aux médias, à savoir identifier intuitivement, dès le début, des indices de type contextuel (type de média, caractéristiques visuelles et sonores, informations relatives à l'instance énonciative...). Ces indices leur permettent de se doter d'une représentation du genre de texte auquel ils sont exposés.

Nous faisons l'hypothèse que ce travail de réflexion, de questionnement et de contextualisation des textes amènera les étudiants vers une compréhension et une appropriation des spécificités du texte. Les compétences langagières acquises en phase de compréhension pourraient ensuite faciliter l'action des étudiants sur le plan langagier, en phase de production.

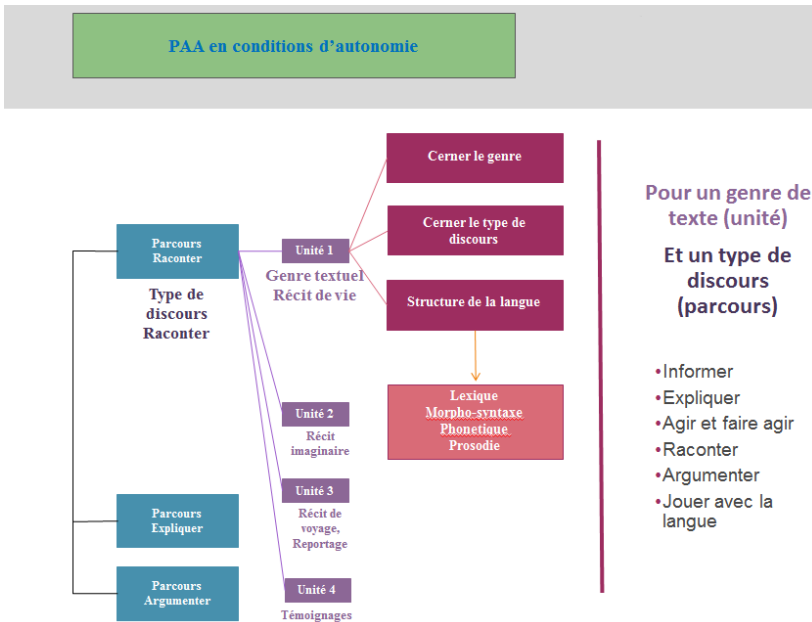
#### 4.2. *La structure des parcours*

Les notions de « genres de texte » et de « type de discours » sont fondatrices de l'architecture globale de nos parcours. En dépit de son caractère composite, le texte que nous sélectionnons et qui devient un *objet didactique*, peut être clairement identifié par sa nature discursive. Ce qui nous intéresse, c'est précisément de pouvoir mettre en exergue la spécificité de cette nature discursive. Nous départageons donc nos ressources en fonction de critères qui nous permettent de les identifier comme appartenant à une des six catégories suivantes : informer, expliquer, agir et faire agir, raconter, argumenter, jouer avec la langue<sup>6</sup>.

A partir de chacune de ces six catégories de types de discours nous construisons un parcours didactique. Chaque parcours se compose de quatre unités. Chaque unité s'élabore sur la base d'un texte (vidéo et authentique) qui relève d'un genre donné. Chaque unité didactique pourra relever (au mieux et si possible, en fonction du type de discours) d'un genre différent.

---

<sup>6</sup> Ces catégories s'inspirent des séquentialités définies par Adam et de la catégorisation des types de discours élaborée par Bronckart (cf. infra, 1.1.) ; elles ont été reformulées afin d'être adaptées à notre contexte d'enseignement.



**Figure 1 - Architecture des PAA**

Ce travail implique le passage d'une vidéo extraite de son contexte d'origine à une vidéo devenue un objet destiné à être enseigné. Les éléments visuels et langagiers sont donc interprétés et repensés en fonction de nos objectifs didactiques.

Nous avons adopté une approche des données de langue qui part de la globalité du texte et s'intéresse ensuite à ses constituants à différents niveaux, adoptant un grain d'analyse de plus en plus fin. La didactisation des ressources fondée sur cette approche vise à développer une compréhension progressive du texte oral à travers un processus qui prévoit dans un premier temps une mise en contexte et une réflexion autour du genre identifié, ensuite l'identification de l'organisation relative au type de discours dominant et seulement à la fin le travail sur la structure de la langue, à savoir lexicale, morphosyntaxe, phonétique, prosodie. Les activités sont de ce fait réunies en trois sections : activités sur le genre, activités sur le type de discours, activités sur le lexique, la morphosyntaxe et la prosodie.

Toutes les activités qui composent l'unité sont proposées entièrement à l'oral<sup>7</sup>. Il s'agit d'un choix didactique qui respecte un des principaux buts des PAA, le travail de perception et de compréhension de l'oral.

<sup>7</sup> Non seulement le texte-vidéo mais aussi toutes les autres composantes de l'activité, notamment : la capacité à mobiliser, les questions, les amorces et les propositions de réponse.

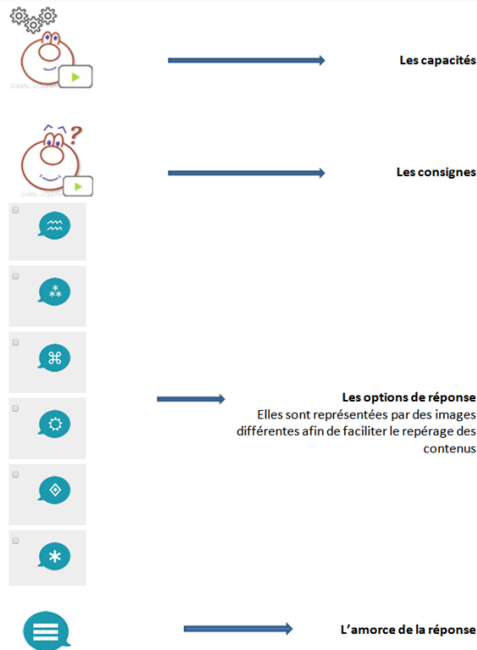


Figure 2 - Le 'tout à l'oral'

Pour chaque activité, nous avons explicité la capacité que nous souhaiterions mobiliser (c'est-à-dire le but de l'activité) pendant la compréhension du texte. Son explicitation est censée permettre à l'apprenant d'identifier plus facilement l'objectif à atteindre, ce qui participe en principe à le rendre plus engagé, voire plus autonome, dans son apprentissage (Masperi & alii, 2013).

### 4.3. Illustration d'une unité prototypique

Afin d'expliciter la façon dont les concepts théoriques et l'approche didactique précédemment illustrés ont été repensés à l'intérieur des PAA, nous proposons l'illustration d'une unité prototypique du PAA d'italien de niveau B2, parcours *Spiegare* (Expliquer). Le texte sélectionné pour cette unité est une vidéo qui actualise le *genre de l'interview*. L'interviewé est le réalisateur italien Michelangelo Frammartino, questionné sur son film *Le quattro volte*. L'interview, intégrée dans les « bonus » du DVD du film, a été montée de façon à montrer uniquement les réponses du réalisateur (les questions de l'intervieweur ne sont donc pas incluses dans la vidéo). Elle est divisée en plusieurs chapitres dont les titres apparaissent à l'écran au fur et à mesure que les différentes séquences tirées du film se succèdent. Une série d'éléments visuels et sonores de type contextuel donnent ainsi des indices sur le genre<sup>8</sup> et sur le type de discours dominant.

<sup>8</sup> Dans le premier chapitre de l'interview, le genre est annoncé à l'écran avec l'élément textuel *Intervista* (Figure 3).





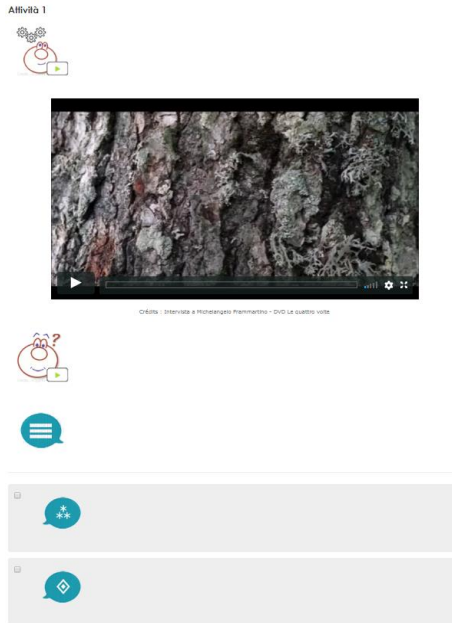
Figure 3 – Captures d’écran du premier chapitre de l’interview *Origini del progetto* (Origines du projet)

#### 4.3.1. Activités d’ancrage dans le genre

L’entrée dans le texte se fait par une série d’activités d’ « ancrage dans le genre ». A partir d’un court extrait, ces activités portent sur l’identification du genre de la vidéo (dans ce cas, l’extrait contient des éléments visuels qui sont essentiels pour pouvoir situer la vidéo à l’intérieur d’un genre). Chaque locuteur possède des connaissances sur le genre dans sa langue maternelle et nous faisons l’hypothèse que c’est sur la base de ces connaissances et intuitions, étayées par des indices de type contextuel, qu’il parvient à identifier le genre auquel il est exposé. Nous envisageons ainsi de créer une *zone de continuité* entre la langue-culture maternelle et la langue-culture cible.

Les activités contenues dans cette section se donnent pour but de guider l’étudiant dans l’identification du genre en passant par le repérage des éléments qui relèvent du contexte, au sens défini plus haut, et qui permettent de faciliter le processus de compréhension globale de la vidéo. Il est important de souligner que les activités du PAA ne sont pas proposées dans une optique d’évaluation. L’un des principaux buts est d’amener l’étudiant à se questionner sur les caractéristiques textuelles et discursives de la vidéo en faisant des hypothèses qu’il pourra vérifier par la suite, grâce à la découverte progressive du texte et aux autres activités proposées.

Dans un premier temps, il est demandé aux étudiants de sélectionner, parmi les options proposées, celles qu’ils considèrent comme étant les plus proches de leurs propres intuitions sur le genre de la vidéo. Il s’agit ici d’un QRM (Questionnaire à Réponse Multiple) dans lequel toutes les options proposées peuvent être considérées comme correctes (Figures 3 et 4).



**Figure 4 – Capture d'écran Activité 1 du PAA B2 Expliquer**

<p><b>Activité 1</b></p> <p>Capacité : Cerner le genre</p> <p>Consigne : Vous allez visionner l'interview du réalisateur Michelangelo Frammartino. Ecoutez les énoncés suivants. Lesquelles de ces propositions sont proches de vos propres intuitions sur les caractéristiques dominantes dans le genre d'interview qui vous est proposée? Vous pourrez les vérifier à la fin du travail.</p> <p>Amorce : Au cours de l'interview, le réalisateur Michelangelo Frammartino va être amené à...</p> <ul style="list-style-type: none"> <li>• expliquer les choix artistiques à la base de son film</li> <li>• raconter des épisodes survenus pendant le tournage</li> <li>• situer le film à l'intérieur de sa filmographie</li> <li>• évoquer la situation du cinéma italien</li> <li>• mettre en valeur ses acteurs</li> <li>• répondre à des questions ponctuelles de l'intervieweur</li> </ul>
---

**Figure 5 - Activité 1 du PAA B2 Expliquer<sup>9</sup>**

Cette première activité est suivie d'une question ouverte qui la complète (Figure 4). Il est en effet proposé aux étudiants d'élargir l'éventail des intuitions avancées sur le genre de la vidéo à travers une réponse ouverte écrite qui sera à l'avenir transposée en réponse ouverte à l'oral (grâce à la fonctionnalité de l'enregistreur offert par l'ENPA, Claroline Connect) afin de respecter le principe du 'tout à l'oral' (cf. infra, 3.2.).

<sup>9</sup> Toutes les activités du PAA de niveau B2 sont proposées à l'oral et en italien.

<p><b>Activité 2</b></p> <p>Capacité : Cerner le genre</p> <p>Consigne : Ajoutez vos propres intuitions sur le genre de la vidéo.</p> <p>Amorce : Au cours de l'interview, le réalisateur Michelangelo Frammartino va être amené à...</p> <hr/> <hr/> <hr/>
---

**Figure 6 - Activité 2 du PAA B2 Expliquer**

#### 4.3.2. *Activités d'ancrage dans le type de discours*

Après avoir contextualisé la vidéo, nous proposons une série d'activités qui fait ressortir l'ensemble des éléments langagiers directement liés au type de discours dominant. Partant du principe que chaque type de discours révèle de manière plus ou moins évidente une organisation particulière des unités de la langue étudiée, nous considérons que cette deuxième section d'activités vise un double objectif : d'une part, se focaliser sur les segments de texte dans lesquels émergent les structures linguistiques qui sont propres au type de discours dominant (dans ce cas le type de discours explicatif) ; et d'autre part, cerner les choix énonciatifs adoptés pour atteindre le but visé. Cela permet d'approcher les conventions culturelles qui dessinent les contours du texte et son adéquation au but énonciatif visé. Autrement dit, il s'agit de décomposer une structure discursive adoptée portée par l'énonciateur et de la soumettre à l'observation de l'étudiant, de manière à alimenter sa réflexion et à développer des capacités langagières susceptibles de l'amener à utiliser potentiellement les structures observées dans des situations communicatives appropriées.

C'est à ce stade de la réflexion que l'outillage numérique et techno-pédagogique a été fortement questionné. Pendant la phase de conception didactique, les activités ont été conçues afin d'activer un certain type d'opérations langagières et il s'avérait que l'outillage numérique offert par l'ENPA ne nous permettait d'atteindre que partiellement nos objectifs. Ce constat nous a amenées à formuler des propositions techno-pédagogiques répondant à nos contraintes et à nos préconisations d'assistance au processus de compréhension de l'oral. L'exemple qu'illustre le mieux les résultats de cette réflexion est le *Chapitrage*, un outil qui a été développé à l'extérieur de l'ENPA mais qui peut s'associer aux protocoles disponibles dans la plateforme (par exemple un QRU ou un QRM, une association, une réponse ouverte etc.)<sup>10</sup>. Le principe de cette activité repose sur le repérage d'éléments langagiers spécifiques : l'étudiant peut sélectionner un segment de texte – dans ce cas un chapitre audio – à partir d'une bande sonore prédécoupée par le concepteur. Le *Chapitrage* nous a ainsi permis de disposer d'un outil qui rend possible la catégorisation des segments de discours qui organisent les passages de textes dans lesquels émergent les structures langagières qui relèvent du type de discours dominant.

Dans l'unité explicative en question, cet outil a été utilisé, par exemple, pour demander d'identifier le segment dans lequel le réalisateur explique « les raisons de la marginalisation des bergers » (cf. Attività 7 - Figure 7). Le texte a été prédécoupé en quatre segments qui peuvent être réécoutés à volonté, y compris de manière non linéaire (cf. boutons 1, 2, 3, 4 – Figure 7). A cette bande sonore a été associé un QRU (question à réponse unique, cf. segmento 1, 2, 3, 4), qui permet d'opérer le bon choix.

<sup>10</sup> Conçu, réalisé et mis gracieusement à disposition par Jean-Jacques Quintin, techno-pédagogue, chercheur en Sciences de l'Éducation à l'Université Lumière, Lyon2.

## Attività 7



**Figure 7 – Chapitrage intégré dans une activité (QRU)**

Nous ne sommes pas encore en mesure de lancer une véritable expérimentation. Nous disposons toutefois des informations issues d'une première expérimentation exploratoire menée au cours de l'année académique 2017-2018, sur un échantillon d'étudiants inscrits aux cours d'italien « Langue et Communication » (LANSAD) de niveau B1.2, B2.1 et B2.2. L'objectif principal était d'observer la conduite des étudiants face aux activités sur le « genre » et le « type de discours » en autonomie totale et de mettre à l'épreuve les hypothèses formulées en phase de conception. Les données recueillies à partir de l'unité présentée laissent apparaître l'expression d'un certain intérêt vis-à-vis des activités d'« ancrage dans le genre ». Ces activités auraient aidé nos informateurs à mieux contextualiser la vidéo proposée. Certains ont ainsi déclaré que les hypothèses à faire au début de l'unité « *permettent de s'interroger sur des choses auxquelles [on n'aurait] pas pensé en visualisant la vidéo une première fois* » ; « *parce que faire une petite réflexion sur le thème et l'hypothèse aide à faire une liaison entre les idées principales et secondaires, pour les séparer aussi de l'information additionnelle* ».

## 5. Conclusion et perspectives

S'approprier une telle approche, théorique et didactique, est un défi de taille. Il s'agit tout d'abord de faire face à des difficultés idiosyncrasiques, qui relèvent de traditions et de pratiques d'enseignement ancrées dans chaque langue-culture. Le plus souvent, les activités de compréhension se focalisent sur le contenu du texte (le « quoi », le « qui », le « où » ?)

plutôt que sur la façon dont le contenu est organisé dans un discours. L'appropriation de la démarche méthodologique présentée dans ce travail nous a permis de formuler une première modélisation de nos parcours d'apprentissage en italien. Les avancées engrangées, issues des réflexions théoriques, du travail d'équipe et de nos expérimentations nous conduisent parallèlement à une première modélisation des parcours en langue anglaise.

Grace à des outils techno-pédagogiques, tel le *Chapitrage*, conçus pour répondre à une attente clairement explicitée sur le plan didactique, nous avons pu conduire les étudiants vers une compréhension plus approfondie, qui se glisse au cœur de la structure du discours. Cela permet notamment de proposer des activités très ciblées non pas sur le résultat de la compréhension mais sur le processus de décodage et de construction du sens. D'autres outils sont actuellement en cours de développement informatique et/ou d'intégration à l'intérieur d'une version enrichie de l'ENPA, Claroline Connect (il s'agit notamment du *Bornage*, de l'*Ecoute Active Enrichie*, de la *Vidéo Active Comparative* et d'un enregistreur son et vidéo). Cela nous permettra d'exploiter plus finement encore les atouts de l'approche par « genre de textes » et par « types de discours ». Le travail de compréhension de l'oral sera proposé à travers des activités de questionnement et d'activation du processus de compréhension de manière plus complète, approfondie et cohérente.

Enfin, une réflexion sur la constitution d'un « corpus » qui permette aux étudiants et aux enseignants de disposer d'une plus grande variété de textes-vidéos est amorcée. Les textes réunis dans le corpus seraient indexés informatiquement afin d'être inscrits dans des pratiques sociales précises, dans un genre et un type de discours dominant. Les usagers –enseignants, concepteurs, étudiants – pourraient ainsi prendre facilement conscience des relations qui existent entre les textes travaillés dans chaque unité des PAA et d'autres textes, ancrés dans la même pratique sociale.

### **Références bibliographiques**

- ADAM, Jean-Michel (1987). Textualité et séquentialité. L'exemple de la description. *Langue française*, 74 (1) : 51-72.
- ADAM, Jean-Michel (1992). *Les textes types et prototypes. Récit, description, argumentation, explication et dialogue*. Paris : Armand Colin.
- ADAM, Jean-Michel (1997). Genres, textes, discours: pour une reconception linguistique du concept de genre. *Revue Belge de Philologie et d'Histoire*, 3/75 : 665–681.
- ADAM, Jean-Michel (2005). *La linguistique textuelle : introduction à l'analyse textuelle des discours*. Paris : Armand Colin.
- ADAM, Jean-Michel (2006). Texte, contexte et discours en questions. Réponses de Jean-Michel Adam. *Pratiques*, (129-130) : 21-34. Consulté à l'adresse : [https://www.unil.ch/fra/files/live/sites/fra/files/shared/Entretien\\_Pratiques-Adam.pdf](https://www.unil.ch/fra/files/live/sites/fra/files/shared/Entretien_Pratiques-Adam.pdf) le 31.01.2019.
- ADAM, Jean-Michel (2011). *La linguistique textuelle. Introduction à l'analyse textuelle des discours*. Paris : Armand Colin.
- ADAM, Jean-Michel (2015). *Faire Texte : Frontières textuelles et opérations de textualisation*. Besançon : Presses Universitaires De Franche-Comté.
- BEMPORAD, Chiara, & PAHUD, Stéphanie (2016). Plurilinguisme, genres et appropriation : une réflexion didactique autour du genre « courrier des lecteurs » en classe de FLE, in *Statut des genres en didactique du français* CORDEIRO, Gláfs Sales (Dir.) & VRYDAGHS, David. Namur : Presses universitaires de Namur, 261-278.
- BIAGIOTTI, Triscia (2017). *La construction d'un parcours d'apprentissage en ligne à travers une approche par « genres de texte » et « type de discours »*. Mémoire de Master 2, sous la direction de : MASPERI, Monica. Université Grenoble Alpes.

- BRONCKART, Jean-Paul (1996). *Activité langagière, textes et discours: pour un interactionisme socio-discursif*. Lausanne : Delachaux et Niestlé. Consulté à l'adresse : <https://archiveouverte.unige.ch/unige:37758> le 31.01.2019.
- BRONCKART, Jean-Paul (1997). La temporalité des discours, comme contribution à la reconfiguration des actions humaines. *Travaux du centre de recherches sémiologiques*, (65) : 39-65.
- BRONCKART, Jean-Paul (2004). Les genres de textes et leur contribution au développement psychologique. *Langages*, 38 (153) : 98-108.
- BRONCKART, Jean-Paul (2008a). Du texte à la langue, et retour. Notes pour une « reconfiguration » de la didactique du français. *Pratiques*, 137/138 : 97-116.
- BRONCKART, Jean-Paul (2008b). Genres de textes, types de discours, et « degrés » de langue, in Gérard, Christophe (Dir.), *Texte, XIII* (1/2). Consulté à l'adresse : <http://www.revue-texto.net/index.php?id=86> le 31.01.2019.
- BRONCKART, Jean-Paul, SCHNEUWLY, Bernard, & REUTER, Yves (2016). *Pourquoi et comment devenir didacticien ?* Villeneuve d'Ascq : Presses universitaires du Septentrion.
- BURGER, Marcel & JACQUIN, Jérôme (2015). La textualisation de l'oral : éléments pour une description de la construction collaborative de la complétude, in *Faire texte* ADAM, Jean-Michel (Dir). Besançon : Presses Universitaires de Franche-Comté, 277-317.
- CANELAS-TREVISI, Sandra (2009). *La grammaire enseignée en classe. Le sens des objets et des manipulations*. Berne : Peter Lang.
- CANELAS-TREVISI, Sandra & BAIN, Daniel (2009). La grammaire scolaire au service de l'enseignement-apprentissage du texte argumentatif ? Analyse critique de quelques pratiques en classe, in *Pratiques d'enseignement grammatical* DOLZ, Joaquim et SIMARD, Claude. Québec : Les presses de l'Université Laval, 155-176.
- CANELAS-TREVISI, Sandra (2015). Les subordonnées relatives et les textes en contexte : une rencontre impossible ? in *Les subordonnées : corpus, acquisition et didactique*, TREVISIOL, Pascale et KAHERAOUI, Malika (dir.). Rennes : Presses Universitaires de Rennes, 139-158.
- CHARTRAND, Suzanne-G. (2008a). *Progression dans l'enseignement du français langue première au secondaire québécois. Répartition des genres textuels, des notions, des stratégies et des procédures à enseigner de la 1ère à la 5ème secondaire*. Québec français, (Hors-série). Consulté à l'adresse [https://www.intranet.fse.ulaval.ca/fichiers/public/prof/pub\\_2374\\_\\_PUB\\_progression.pdf](https://www.intranet.fse.ulaval.ca/fichiers/public/prof/pub_2374__PUB_progression.pdf) le 31.01.2019.
- CHARTRAND, Suzanne-G. (2008b, novembre). *Travailler les textes en classes, oui, mais par genre*. Congrès de l'AQPF. Consulté à l'adresse [http://www.enseignementdufrancais.fse.ulaval.ca/fichiers/site\\_ens\\_francais/modules/document\\_section\\_fichier/fichier\\_\\_f85b9ce631b9\\_\\_enseigner\\_les\\_textes\\_par\\_les\\_genres\\_fin.pdf](http://www.enseignementdufrancais.fse.ulaval.ca/fichiers/site_ens_francais/modules/document_section_fichier/fichier__f85b9ce631b9__enseigner_les_textes_par_les_genres_fin.pdf) le 31.01.2019.
- CHARTRAND, Suzanne-G. (2008c). Répartition des genres textuels, des notions, des stratégies et des procédures à enseigner de la 1re à la 5e secondaire. *Québec français, numéro spécial*.
- CHARTRAND, Suzanne-G., ÉMERY-BRUNEAU, Judith & SENECHAL, Kathleen avec la collaboration de RIVERIN, Pascal (2015). *Caractéristiques de 50 genres pour développer les compétences langagières en français*. Québec : Didactica.
- CHARTRAND, Suzanne-G. (2016). Les genres du discours : point nodal de la discipline français, in *Statut des genres en didactique du français* CORDEIRO, Gláís Sales (Dir.) & VRYDAGHS, David. Namur : Presses universitaires de Namur, 53-81.
- CLAUDEL, Chantal & LAURENS, Véronique (2016). *Le genre discursif comme objet d'enseignement en didactique du français*. Congrès Mondial de Linguistique Française - CMLF 2016.
- CONSEIL DE L'EUROPE (2001). CECRL (Cadre européen Commun de Référence pour les Langues). Strasbourg : Didier.

- DOLZ, Joaquim, NOVERRAZ, Michèle & SCHNEUWLY, Bernard (Ed.) (2001). *S'exprimer en français. Séquences didactiques pour l'oral et pour l'écrit*. Bruxelles : De Boeck.
- MASPERI, Monica (2012). Innovation et transformation des pratiques de l'enseignement-apprentissage des langues dans l'enseignement supérieur. [Rapport de recherche] UGA – Université Grenoble Alpes. <hal-02000901>
- MASPERI, Monica (2013). Apprendre à s'autoformer en langues : approches créatives et outils numériques, (MASPERI, Monica, Ed.) *Les langues modernes, 4-2013*.
- MASPERI, Monica, & QUINTIN, Jean-Jacques (2014a). L'innovation selon Innovalangues. *LEND (Lingua E Nuova Didattica)* : 6-14.
- MASPERI, Monica, & QUINTIN, Jean-Jacques (2014b). Enseigner à l'université en France, à l'ère du numérique : l'apport de dispositifs innovants dans la formation en langues, in *Dispositivi formativi e modalità ibride per l'apprendimento linguistico*, CERVINI, Cristiana & VALDIVIESO, Anabel. Bologna : Edizioni Clueb, 61-80.
- MIRANDA, Florencia, COUTINHO, Maria Antónia (2015). Les propriétés différentielles des genres et leurs implications didactiques. *Recherches et applications – Le français dans le monde*, 58 : 17- 26.
- MONTE, Michèle, GILLES, Philippe & ADAM, Jean-Michel (2014). *Genres & Textes : déterminations, évolutions, confrontations*. Lyon : Presses Universitaires de Lyon.
- QUARANTA, Giada (2017). *Aspetti della costruzione didattica e approccio per generi nei Percorsi d'Apprentissage en Autonomie nel quadro del progetto Innovalangues*, Mémoire de Master 2 sous la direction de : CERVINI, Cristiana, MASPERI, Monica, Università Alma Mater Studiorum Bologna.
- RASTIER, François (2001). *Arts et sciences du texte*. Paris : PUF.
- SCHNEUWLY, Bernard, DOLZ, Joaquim (1997). Les genres scolaires. Des pratiques langagières aux objets d'enseignement. *Repères*, 15(1) : 27-40. Consulté à l'adresse : [https://www.persee.fr/doc/AsPDF/reper\\_1157-1330\\_1997\\_num\\_15\\_1\\_2209.pdf](https://www.persee.fr/doc/AsPDF/reper_1157-1330_1997_num_15_1_2209.pdf) le 30.01.2019.
- SCHNEUWLY, Bernard (2000). Les outils de l'enseignant – Un essai didactique. *Repères*, 22 : 19-38. Consulté à l'adresse : [https://www.persee.fr/doc/AsPDF/reper\\_1157-1330\\_2000\\_num\\_22\\_1\\_2341.pdf](https://www.persee.fr/doc/AsPDF/reper_1157-1330_2000_num_22_1_2341.pdf) le 30.01.2019.
- SCHNEUWLY, Bernard, & DOLZ, Joaquim (2009). *Des objets enseignés en classe de français – Le travail de l'enseignant sur la rédaction de textes argumentatifs et sur la subordonnée relative*. Rennes : Presses Universitaires de Rennes.
- SCHNEUWLY, Bernard, AEBY DAGHE, Sandrine (2016). Genres scolaires et genres scolarisés en écriture et pratiques sociales de référence. Le cas de l'école primaire en suisse romande (1830-2010). *Recherches*, no. 65 : 163-175.
- THEVENAZ-CHRISTEN, Thérèse & CANELAS-TREVISI, Sandra (2019). L'écriture et la lecture littéraire : une relation qui se transforme, in *Lire des textes réputés littéraires : disciplinazione et sédimentation. Enquête au fil des degrés scolaires en Suisse romande*, RONVEAUX, Christophe et SCHNEUWLY, Bernard (dir.). Bruxelles : Peter Lang, Théocrit', Vol. 13, 305-324.

**LE NUMERIQUE DANS LES METHODOLOGIES DE  
RECHERCHE**



# **ANALYSE DE LA COMBINATOIRE DES VERBES D’AFFECT EN ARABE ET EN FRANÇAIS FONDEE SUR LES CORPUS NUMERIQUES *EMOLEX* ET *ARABICORPUS***

Safa ZOUAIDI<sup>a,b</sup>

[szouaidi@yahoo.com](mailto:szouaidi@yahoo.com)

<sup>a</sup>LLTA, Faculté des lettres et sciences humaines de Sfax, Sfax, Tunisie

<sup>b</sup>LIDILEM, Université Grenoble Alpes, Grenoble, France

## **1. Introduction**

Notre étude s’insère dans le cadre de la linguistique des corpus (Sinclair, 1991). La question principale concerne les correspondances entre le français et l’arabe dans l’expression de l’affect appartenant aux champs sémantiques de la surprise, de la colère, de l’admiration et de la jalousie. Pour les verbes d’affect (V\_affected), certains auteurs en proposent un classement en s’appuyant sur des critères syntaxiques (Harris, 1988 ; Gross, 1995 ; etc.). D’autres proposent une répartition des verbes en se fondant sur des traits sémantiques homogènes (Mathieu, 2000 ; etc.). Nous proposons, ici, de relier systématiquement la sémantique (les dimensions sémantiques véhiculées par les collocatifs) à la syntaxe (les constructions grammaticales récurrentes). Pour cela, nous solliciterons deux corpus journalistiques numériques<sup>1</sup> comparables et équilibrés : *EmoBase* (projet *Emolex* 120 M de mots) et *ArabiCorpus* (130 M de mots).

Ces corpus nous permettront de nous interroger, en particulier, sur les correspondances syntaxico-sémantiques dans les combinatoires verbales entre les deux langues. Nous chercherons, plus précisément, à savoir si des verbes appartenant aux mêmes types d’affect dans les deux langues attireraient les mêmes accompagnateurs ou non.

Pour ce faire, après une présentation du cadre théorique, de l’approche contrastive et de la méthodologie du travail, nous présenterons les corpus sur lesquels cette recherche a été fondée. Nous nous arrêterons, ainsi, sur la constitution des corpus de travail et de leur rapprochement tout en abordant les avantages et les limites. Notre travail analysera, enfin, les combinaisons des verbes d’affect sélectionnés d’un point de vue sémantico-syntaxique, et ce dans les deux langues comparées.

## **2. Les V\_affected : Cadre théorique et méthodologie**

Les V\_affected sont difficiles à classer à cause de leur polysémie (Bouchard, 1995). Il existe aussi des flottements terminologiques : *verbes psychologiques*, *de sentiment*, *affects*, *émotions*, *états affectifs*, etc. Les différents travaux cherchant à circonscrire la classe des verbes d’affect, pour le français, prennent, à notre connaissance, peu en compte le fait de relier systématiquement la sémantique à la syntaxe (Harris, 1988 ; Gross, 1995 ; Mathieu, 2000 ; etc.). Nous remarquons ici l’absence d’analyses reliant systématiquement les dimensions sémantiques véhiculées par les collocatifs aux constructions grammaticales récurrentes.

En ce qui concerne l’arabe, les difficultés sont d’une autre nature. Les travaux dans ce domaine sont beaucoup moins nombreux et il n’existe pas, à notre connaissance, d’études sémantico-syntaxiques des verbes d’affect. Nous avons trouvé quelques remarques, certes, mais nous considérons qu’elles sont trop parcellaires et insuffisantes pour servir d’appui à notre étude (Donat, 1892). La notion même de collocation, est traitée jusque là, autour du couple anglais-arabe sous un angle strictement lexicographique (Abu Ssaydeh, 2001). Ceci constitue, certes,

<sup>1</sup> Nous n’avons pas trouvé de corpus comparables arabe-français alignés appartenant à la même époque.

une difficulté, mais justifie pleinement la recherche entreprise qui est, dans ce sens, novatrice. Dans notre travail, le terme d'affect est utilisé non pas comme une sous-classe des affects, mais comme une étiquette conventionnelle générique<sup>2</sup>. Quant au terme *collocation*, nous l'utilisons au sens de Hausmann (1989, 2007), c'est-à-dire comme une combinaison phraséologique binaire composée d'une base et d'un collocatif.

Par ailleurs, dans l'objectif de proposer une analyse de la combinatoire des verbes d'affect dans les deux langues, sur le plan sémantique et syntaxique, nous nous sommes principalement appuyée sur la théorie du *Lexical priming* de Hoey (2005). Cette théorie fait partie des thèses contextualistes et s'intéresse à l'emploi d'un mot dans des associations sémantiques que chaque locuteur peut amorcer en s'appuyant sur son savoir linguistique<sup>3</sup>. Il s'agit, ici, d'étudier le degré d'attirance entre les lexies en examinant les « préférences » (*admirer secrètement*) ou les « aversions » (*\*étonner secrètement*) entre les mots et d'appréhender, ainsi, pourquoi certaines « colligations<sup>4</sup> » peuvent attirer des lexies particulières et non pas d'autres, bien qu'elles soient quasi synonymes.

Notre analyse se focalise sur les champs de la surprise, de la colère, de la jalousie et de l'admiration. Ce choix représente une palette suffisamment variée des types d'affect, comme le montre le tableau suivant :

Champs	V_pivot en français	Équivalent V_pivot en arabe
<b>V_émotion (réactifs)</b> deux actants (X/Z)	Surprise	étonner أدهش [[ʔadhaʃa]
	Colère	énervé أغضب [[ʔayɖaba]
<b>V_sentiment (interpersonnels)</b> deux actants (X/Y)	Jalousie	envier حسد [hasada]
	Admiration	admirer أعجب [[ʔaʕʒaba]

**Tableau 1 – Les verbes pivots sélectionnés en français et en arabe**

Comme l'indique le tableau ci-dessus, nous avons deux verbes qui renvoient à des affects causés, *étonner* et *énervé*, appelés *émotions* (l'un de polarité neutre, l'autre de polarité négative). Les deux autres verbes *admirer* et *envier*, renvoient à des affects interpersonnels (appelés *sentiment*), respectivement de polarité positive et négative. Nous nous sommes beaucoup appuyée sur les apports théoriques et méthodologiques développés dans le cadre du projet Emolex pour l'analyse du lexique affectif qui repose sur l'étude contrastive de la combinatoire syntaxique et lexicale des lexies d'affect (Novakova et Tutin, 2009).

Dans la section suivante, nous présenterons nos deux corpus.

<sup>2</sup> L'affect recouvre ici des termes relevant à la fois des émotions et des sentiments, même si, au cours de l'étude, nous serons amenée à bien distinguer ces deux sous-catégories, avec les V\_émotion (causés, réactifs et ponctuels) et les V\_sentiment (interpersonnels et duratifs) (cf. Ruwet, 1995 ; Mathieu, 2000).

<sup>3</sup> Les principes de la théorie de Hoey (2005) sont appliqués, jusque-là, à notre connaissance, seulement en anglais et à un nombre limité de lexies, ce qui motive l'étude entreprise.

<sup>4</sup> Les environnements grammaticaux portent l'étiquette de « colligations » chez Hoey (2005).

### 3. Présentation des corpus

Il s'agit, en effet, de deux corpus journalistiques numériques comparables et équilibrés : *EmoBase* (projet *Emolex*<sup>5</sup> 120 M de mots) et *ArabiCorpus* (130 M de mots).

L'interface *EmoBase* est interrogeable en ligne (Kraif, 2002) et permet d'obtenir des résultats détaillés concernant toutes les occurrences d'un pivot donné dans le corpus. Quant à l'interface *ArabiCorpus*<sup>6</sup>, elle est moins ergonomique que celle d'*Emolex*. Elle ne dispose pas, en effet, de structure d'annotation ni, d'ailleurs, d'indexation préstructurée, ce qui peut rendre l'opération de repérages des collocations en question beaucoup plus chronophage. Ces corpus nous permettent de nous interroger, en particulier, sur les correspondances syntaxico-sémantiques dans les combinatoires verbales entre les deux langues.

Tout d'abord, pour le français, nous avons rassemblé les lexicogrammes, qui constituent des « tables de contingences » (Diwersy *et al.*, 2014) réalisées sur la base du calcul de *log-likelihood* et qui permettent la classification des combinaisons verbales retenues selon des paramètres syntaxiques et sémantiques prédéfinis, comme l'illustre l'extrait suivant :

The screenshot shows the Lexicoscope web interface. At the top, it says 'Lexicoscope Concordances'. Below that, there's a search bar and a list of concordances. The table below is an extract of the results shown in the interface.

Num	Identifiant	Contexte gauche	Pivot	Contexte droit
241	s300170	Les jeunes s'	étonnent	de connaître la satiété en mangeant moins et de ne plus avoir faim entre les repas.
242	s301653	Le durcissement du conflit autour de la réforme des ports n'	étonne	guère à Marseille.
243	s302231	Les métamorphoses de sa ville l'	étonnent	encore : « Le soir, avec les enseignes lumineuses des hôtels et des restaurants, on se croirait maintenant à Monte-Carlo &quot; ; dit-il.
244	s302965	A ceux qui s'	étonnent	que la ségrégation raciale sévisse encore dans les églises américaines, quarante ans après la mort de Martin Luther King, le chroniqueur du Washington Post Colbert King rappelait récemment l'histoire de Richard Allen et d'Absalom Jones, les premiers Noirs ordonnés pasteurs aux États-Unis, expulsés par la congrégation qui les avait acceptés.
245	s304359	A Elista, la rue Lénine offre un spectacle plus réjouissant et, surtout,	étonnant	: des petits kiosques en forme de pagode vendent de l'encens ou des portraits du dalai-lama, tandis qu'un bouddha jaune trône au

Tableau 2 – Extrait des résultats obtenus dans EmoConc

A partir de l'interface d'interrogation *EmoConc* et *EmoLing*, on peut choisir le pivot en cliquant sur *lemme* pour demander toutes les formes possibles du mot pivot *étonner*, par exemple. Ensuite, on peut choisir le type du patron syntaxique dans lequel se trouve notre pivot, à savoir Verbe-circonstant. On choisit aussi un seuil de fréquence de 5, et on commence l'opération d'extraction en concordances, ce qui permet d'obtenir des profils combinatoires (par exemple : *étonner\_adverbe*, *V\_étonner*, etc.). Dans notre cas, nous avons suivi cette démarche pour toutes les opérations d'extraction des occurrences.

Ensuite, l'interface *ArabiCorpus* permet de réaliser des requêtes pertinentes en partant des lemmes. Pour la langue arabe, nous n'avons pas eu les outils permettant d'obtenir des statistiques liées au *log-likelihood*, comme c'était le cas pour le français. Pour résoudre cette

<sup>5</sup> Le projet franco-allemand ANR/DFG Emolex (ANR-09-FASHS-017), sous la responsabilité d'I. Novakova (Université Stendhal, Grenoble Alpes) et de P. Blumenthal (Université de Cologne), a pour objectif l'étude du lexique des émotions dans cinq langues européennes dans ses aspects sémantique, syntaxique et discursif. <http://emolex.ugrenoble3.fr/emoBase/>.

<sup>6</sup> *ArabiCorpus*, est disponible en ligne <http://arabicorpus.byu.edu/>, avec des fonctionnalités en anglais et réalisée aux États-Unis.

contrainte, il a fallu passer par un tri manuel. Nous avons donc constitué des tableaux Excel qui se rapprochent du modèle des lexicogrammes fait pour le français tout en appliquant le même format de présentation des données, comme le montre l'extrait ci-dessous pour le verbe [ʔaʕzaba] (admirer) :

Tableau 3 – Extrait des résultats obtenus dans *Arabicorpus*

Les résultats de ces calculs correspondent à des résultats en pourcentages et non pas au degré de spécificité *log-likelihood*. Cela nous permet de percevoir l'importance de la fréquence d'une combinaison particulière par rapport à l'ensemble des dimensions véhiculées par les autres combinaisons.

Le tableau 4 suivant dresse la fréquence des verbes pivots sélectionnés dans les deux langues dans les deux corpus :

Champs	V_pivot en français	Équivalent V_pivot en arabe	Nombre d'occurrences	
			Français	Arabe
Surprise	étonner	أدهش [[ʔadhaʕa]	130	90
Colère	énervé	أغضب [ʔaʕdaba]	110	102
Jalousie	envier	حسد [hasada]	17	24
Admiration	admirer	أعجب [ʔaʕzaba]	145	199

Tableau 4 – Nombre d'occurrences des verbes pivots sélectionnés en français et en arabe<sup>7</sup> dans les corpus

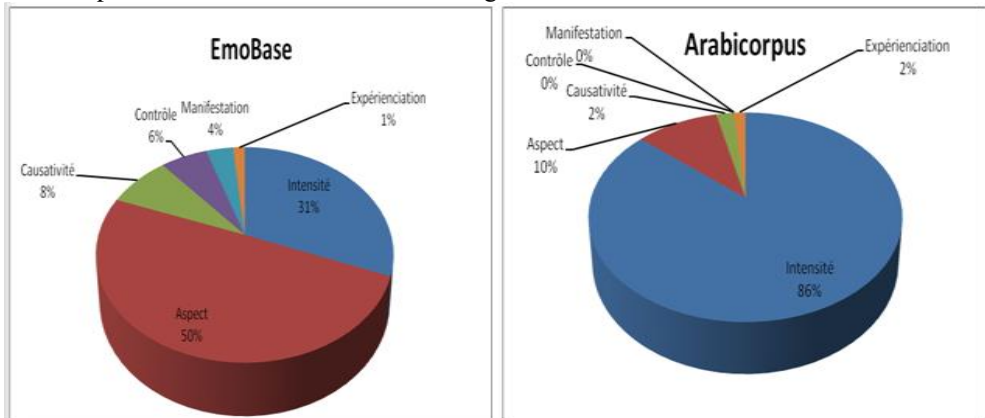
<sup>7</sup> Afin de rendre la lecture des lexèmes arabes plus commode, notamment pour les locuteurs non natifs, nous proposons de les présenter en trois lignes : d'abord, nous les translitérons conformément à l'alphabet phonétique international (API) (Orig.translit.), puis nous proposons une traduction en français du mot original en arabe en vue de gloser les associations verbales concernées selon les règles du *Leipzig Glossing* (Trad.litt.). Enfin, la troisième ligne correspond à la traduction française des exemples arabes et permet de mettre en œuvre la comparaison (Trad.fr.).

Nous avons obtenu au total 817 cooccurrences (402 cooccurrences, en français, et 415 cooccurrences en arabe).

Dans la section suivante, nous proposerons une étude des relations syntaxiques privilégiées qui correspondent aux dimensions sémantiques que véhiculent les co-occurrences verbales des affects, et ce dans les deux corpus.

#### 4. Analyse sémantico-syntaxique

Selon les données observées, les accompagnateurs, statistiquement spécifiques des V\_ affect renseignent sur différentes dimensions sémantiques véhiculées par les collocations, comme l'*intensité*<sup>8</sup>, l'*aspect*<sup>9</sup>, la *polarité*<sup>10</sup>, la *causativité*<sup>11</sup>, la *manifestation*<sup>12</sup>, le *contrôle*<sup>13</sup>, l'*expérimentation*<sup>14</sup>. La répartition des dimensions sémantiques n'est pas équilibrée dans les deux corpus comme l'illustrent les deux émogrammes suivants :



**Figure 1 – Émogrammes avec la répartition des dimensions sémantiques véhiculées par la combinatoire des V\_affect dans les corpus EmoBase et ArabiCorpus**

Comme nous pouvons le constater, nous pouvons remarquer une prédominance des collocatifs des V\_affect véhiculant l'*intensité* (86 %) dans le corpus arabe. Les trois autres dimensions, en l'occurrence l'*aspect* (10 %), la *causativité* (2,17 %) et l'*expérimentation* (1,44 %), sont presque également réparties et assez marginales. Inversement, la dimension aspectuelle est plus marquante en français qu'en arabe. Ceci pourrait s'expliquer par la spécificité de la langue arabe qui, contrairement, au français, morphologise l'aspect. Le français, en revanche, n'a pas de morphème et a besoin davantage des accompagnateurs aspectuels.

<sup>8</sup> L'intensité regroupe les collocatifs référant au degré d'intensité de l'affect (*s'énerver prodigieusement, étonner particulièrement*).

<sup>9</sup> La dimension de l'aspect regroupe les collocatifs qui expriment la représentation du déroulement d'un affect dans le temps (ex. étonner encore, admirer toujours, commencer à admirer, finir par s'énerver, etc.)

<sup>10</sup> La polarité regroupe les collocatifs indiquant l'évaluation de l'affect sur deux axes : négative ou positive.

<sup>11</sup> La causativité regroupe les collocatifs qui expriment le fait qu'un événement causateur, une personne ou une chose amène quelqu'un à ressentir un affect (ex. *faire admirer*, etc.).

<sup>12</sup> La *manifestation* regroupe des modificateurs qui expriment la façon dont certains affects sont extériorisés par l'expérimenteur et peuvent être perçus par un tiers (ex. *sembler (s')étonner, sembler admirer, (s')énerver visiblement*, etc.).

<sup>13</sup> La dimension « contrôle » réunit les collocatifs dénotant l'(in)capacité de l'expérimenteur à contrôler l'affect (ex. *se laisser s'énerver, s'énerver facilement*, etc.)

<sup>14</sup> La dimension de l'expérimentation réunit les collocatifs exprimant, sans spécification supplémentaire, qu'une personne expérimente un affect (ex. *s'énerver presque*).

Dans les deux sous-sections suivantes, nous nous proposerons d'observer respectivement les paramètres intensif et aspectuel en corpus.

#### 4.1. *La dimension intensive*

Pour la dimension de l'intensité, la comparaison entre les deux corpus français et arabe montre des structures différentes entre les deux langues. Les combinaisons relevées, malgré leurs spécificités, répondent, en effet, à des configurations syntaxiques binaires, notamment en français, où la structure récurrente est (V + Adv) (ex. *énervé prodigieusement*<sup>15</sup>).

Du point de vue contrastif, le champ des V\_ affect s'organise de manière très différente, en arabe :

- (V + Adj<sup>accusatif</sup>) ([ʔadhafa kaθīran] (étonner dense)) (41 %),
- (V + N<sup>accusatif</sup>) ([ʔaʕzaba ʕiddan] (admirer sérieux)) (30 %),
- (V + N + N) ([ʔaʕzaba ʔaʕadda ʔal-ʔiʕzābi] (admirer sévérité admiration)) (5%),
- (V + Prép + N), ([ʔadhafa bi-ʕumqin] ([il]étonne avec-profondeur)) (11%).

Il y existe également des structures ternaires (exclusivement pour la dimension de l'*intensité*) : (V + N + Adj<sup>accusatif</sup>) (ex. [ʔaʕzaba ʔiʕzāban kabīran] « admirer admiration grande ») (13% de l'ensemble des configurations syntaxiques relevées dans le corpus).

Ainsi, dans la mesure où les structures tripartites relevées peuvent constituer une superposition de deux collocations, notre analyse pourrait remettre en cause le statut strictement binaire des collocations (Hausmann, 1989, 2007). L'arabe ne structure donc pas la dimension sémantique intensive des affects étudiés de la même façon que le français. En français, il s'agit exclusivement de modificateurs adverbiaux, tandis que l'arabe manifeste plus de diversité au niveau formel des associations lexicales.

Nous passons, dans la sous-section suivante, à l'étude des particularités des associations qui expriment la dimension *aspectuelle*.

#### 4.2. *La dimension aspectuelle*

Les verbes d'affect se combinent différemment en fonction du type d'affect auquel ils appartient. La figure suivante présente la répartition des associations des V\_ affect pour les valeurs ponctuelle et durative en français :

<sup>15</sup> En français, nous pouvons trouver des occurrences de type V+Prép+N, par exemple « énerver avec insistance », « énerver sans scrupule », « admirer sans retenue », etc. Cependant, ces constructions sont très peu fréquentes dans notre corpus et nous ne retenons que celles qui sont statistiquement fréquentes.

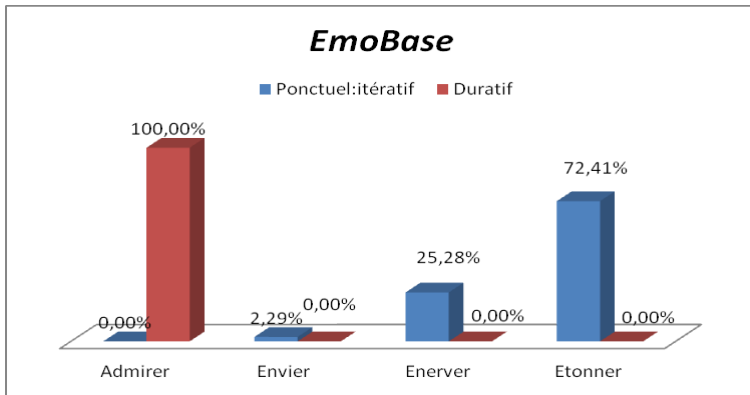


Figure 2 – Valeurs ponctuelle vs durative en français (en %)

Comme l'illustre la figure (2), les quatre V\_affect s'associent à des modificateurs aspectuels, mais dans des proportions variables. La combinatoire des V\_émotion *étonner* et *énervé* ne véhiculent pas l'aspect duratif : elle a tendance à exprimer préférentiellement le ponctuel itératif (*encore, parfois*). La combinatoire du V\_sentiment *admirer* exclut la valeur ponctuelle itérative et exprime davantage la valeur durative (100 %). L'association avec le V\_envier, quant à elle, est absente des fréquences pour la valeur durative. Elle présente tout de même une préférence pour l'aspect ponctuel itératif (2,29 %). Celui-ci renvoie à des affects qui sont de courte durée et qui se répètent.

Du point de vue contrastif, la figure suivante présente la fréquence des associations verbales d'affect qui véhiculent les mêmes valeurs (ponctuelle itérative vs durative) en arabe :

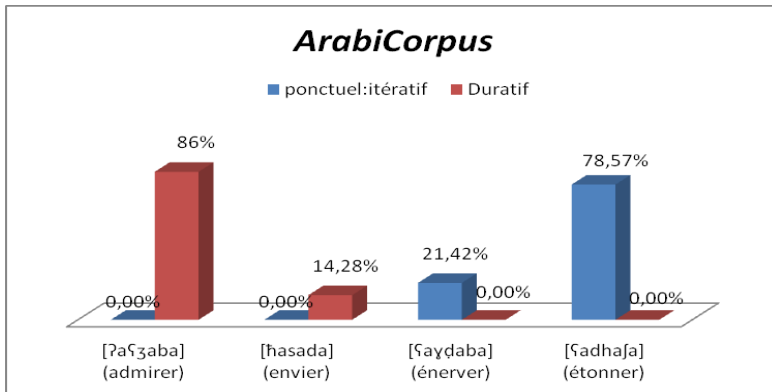


Figure 3 – Valeurs ponctuelle itérative vs durative en arabe (en %)

On y observe la préférence qu'ont les V\_émotion [ʔaʕdaba] (énervé) et [ʕadhafa] (étonner) à se combiner avec des collocatifs ponctuels. C'est une différence avec les V\_sentiments [ʔaʕʕaba] (admirer) et [hasada] (envier), qui ont tendance à s'associer davantage avec des modificateurs duratifs.

Partant de ces observations, pour la dimension aspect, la combinatoire des V\_émotion exprime davantage la valeur ponctuelle itérative, et ce, pour les deux verbes ((s')étonner/[ʕadhafa] et (s')énervé/[ʕaʕdaba]) en français et en arabe. En revanche, les V\_sentiment véhiculent davantage la valeur durative :

- **V\_émotion** *étonner*/[ʕadhafa] et *énervé*/[ʕaʕdaba] + **collocatifs ponctuels** (*étonner encore*, [ʕadhafa ʔahjānan] (étonne-[il] instants)).

- **V\_sentiment admirer**/[ʔaʕʒaba] et **envier**/[ħasada] + **collocatifs duratifs** (*admirer toujours*, [ʔaʕʒaba dāʔiman] (il-admire qui-dure)).

L'analyse comparative nous permet de constater également que ces valeurs ne sont pas régulières. Par exemple, le même collocatif *toujours*, étant sémiotiquement dépendant du V *admirer*, indique que l'expérienceur poursuit dans le sentiment ressenti et correspond, ainsi, à la notion d'éprouver un sentiment duratif :

- (1) J'ai **toujours admiré** sa voix divine, au timbre si reconnaissable. (Valeur durative)

En revanche, l'étude de ce modifieur en contexte montre qu'il peut pencher plutôt du côté de l'aspect ponctuel<sup>16</sup> tout en acceptant en même temps des emplois duratifs, qui indiquent que l'expérienceur ressent l'émotion à chaque fois. Voici les deux exemples suivants :

- (2) Nicolas Sarkozy **s'étonne toujours** de l'image que la société lui renvoie de lui-même : « On me dit que je fais peur, dit-il, on me prend pour un monstre. » (Valeur ponctuelle itérative)

- (3) **Orig.AR** : لِيَدِهِنَّ دَائِمًا بَطْلٌ مِثْلُ الْمَاشِي تَلْقَفُ حِكَايَتَهُ رَوَائِي مَبْدَعٌ فَقْرًا الْحَيَاةَ (Thawra)

**Orig.translit** : [li-judhiʕa-nā dāʔiman baʕalun miθlu ʔal-māʕī talaqqafa ħikājata-hu riwāʔijjun mubdiʕun fa-qaraʔa ʔal-ħayāt]

V transitif direct-1<sup>ère</sup> pers pluriel présent N accusatif

**Trad.litt.** : (pour-étonne-nous qui-dure héro comme almachi a-ramassé histoire-lui romancier doué alors-a-lu la-vie)

**Trad.fr.** : Ça nous **étonne toujours**, un champion comme Almachi dont l'histoire a été saisie par un romancier artiste qui a compris la vie.

Dans les exemples ci-dessus, il s'agit d'une reprise ponctuelle qui marque que le locuteur ne ressent pas l'étonnement durant la totalité du temps, mais seulement à chaque fois que la société décrit son profil, en [2]. En [3], le modifieur ponctuel et duratif [dāʔiman] (qui-dure) est un nom d'agent accusatif qui peut être rendu en français par l'Adv « toujours » ; il exprime tout comme *encore* l'aspect itératif des émotions ponctuelles (surprise). Il peut signifier à la fois « instant » et « qui-dure », et est ainsi capable d'exprimer cette valeur aspectuelle ponctuelle durative dans un contexte donné en combinaison avec un V\_émotion.

C'est en l'occurrence l'aspect télique du verbe « s'étonner » qui conditionne l'interprétation itérative (cf. Gosselin, 1996<sup>17</sup>).

L'étude du degré d'attrance ou d'aversion entre les V\_affect et leurs accompagnateurs permet de mieux appréhender les liens qui font que certains de ces verbes peuvent attirer des collocatifs particuliers et non pas d'autres.

## 5. Conclusion

L'étude sur corpus montre que les combinaisons verbales d'affects étudiées ici véhiculent des dimensions sémantiques qui sont inhérentes aux structures syntaxiques spécifiques.

<sup>16</sup> Dans l'exemple suivant, issu du *Grand Robert* : *Il arrive toujours à cinq heures*, nous n'avons aucune idée de durée, mais de reprise ponctuelle du procès. Et aussi dans un autre exemple du *Grand Robert* : *Il est en retard, lui toujours ponctuel*, l'adverbe *toujours* signifie qu'à chaque fois il est ponctuel.

<sup>17</sup> Dans Gosselin (1996), les différents éléments linguistiques peuvent créer parfois des conflits qui nécessitent des stratégies spécifiques pour les résoudre. Par exemple, « Pierre mangeait sa soupe en cinq minutes » donne lieu à un « conflit » créé entre l'imparfait (*mangeait*) et le complément de durée (*en cinq minutes*), qui nous amène à en faire une lecture itérative. Dans notre cas, la télicité est une propriété du verbe « étonner ».



Il s'agit d'une étude qui prend en considérations les interactions entre les niveaux sémantique et syntaxique pour l'analyse du lexique des affects et qui, de ce fait, loin d'être une analyse interprétative nouvelle foncièrement différente des analyses existantes, permet une description d'un certain nombre de phénomènes linguistiques. Nous avons appliqué et développé des analyses d'*Emolex*, mais il ne s'agit pas, bien entendu, d'appliquer de manière systématique et mécanique des outils d'analyse conçus parfois pour une langue spécifique, ce qui constitue un apport de ce travail. L'analyse de la combinatoire des verbes d'affect et la comparaison entre les langues fondées sur des corpus numériques font apparaître des particularités linguistiques qui ne ressortiraient pas nécessairement d'une analyse unilingue.

Des perspectives de recherche peuvent être envisagées. Par exemple, l'étude fondée sur ces corpus numériques pourrait être étendue à d'autres catégories grammaticales comme les adjectifs, les noms ou les adverbes et ce, dans une perspective contrastive. Une nouvelle perspective se dessine également grâce à la prise en compte plus fine de l'approche fonctionnelle (sémantique, syntaxe et discours), ce qui pourrait produire des résultats pertinents.

### ***Références bibliographiques***

- ABU SSAYDEH, Abdul-Fattah. (2001). Synonymy, Collocation and the Translation, *Turjuman*, 10 : 53-71.
- BLUMENTHAL, Peter & HAUSMANN, Franz Josef (DIR.) (2006). Collocations, corpus, dictionnaires, *Langue française*, 150.
- BOUCHARD, Denis (1995). Les verbes psychologiques, *Langue française*, 105 : 6-16.
- DONAT, Vernier S.J. (1892). *Grammaire arabe composée d'après les sources primitives*. t. 2. Beyrouth : Imp. Catholique.
- DIWERSY Sascha, Goossens Vannina, Grutschus Anke, Kern Beate, Kraif Olivier, Melnikova Elena et Novakova Iva (2014). Traitement des lexies d'émotion dans les corpus et les applications d'*EmoBase, Corpus*, 13, 269-293.
- GOSSELIN, Laurent (2016). Sémantique de la temporalité en français: Un modèle calculatoire et cognitif du temps et de l'aspect. Duculot (Collection Champs Linguistiques), Louvain-la-Neuve.
- GROSS, Maurice (1995). Une grammaire locale de l'expression des sentiments, *Grammaire des sentiments, Langue française*, 105 : 70-87.
- HARRIS, Zellig Sabbeta (1988). *Language and Information*. New York : Columbia University Press.
- HAUSMANN, Franz Josef (1989). Le dictionnaire de collocations, in F.J. Hausmann, O. Reichmann, H.E. Wiegand et L. Zgusta (eds), *Wörterbücher: ein internationale Handbuch zur Lexicographie. Dictionnaires, Dictionnaires*, Berlin/New-York, De Gruyter, 1010-1019.
- HAUSMANN, Franz Josef (2007). Lexicologie française et phraséologie, *Collocation, phraséologie, lexicographie*, Aachen, Shaker Verlag, 121-53.
- HOEY, Michael (2005). *Lexical priming: a new theory of words and language*. London/New York : Routledge.
- KRAIF, Olivier (2002). Méthodes de filtrage pour l'extraction d'un lexique bilingue à partir d'un corpus aligné, *Lexicometrica, Alignement lexical dans les corpus multilingues*, 1-22.
- LE GRAND ROBERT, 9<sup>e</sup> édition, Version informatisée (dernière consultation le 05.01.2019).
- MATHIEU Yvette Yannick (2000). *Les verbes de sentiment : de l'analyse au traitement automatique*. Paris : Éditions du CNRS.
- NOVAKOVA, Iva & TUTIN, Agnès (éds) (2009). *Le lexique des émotions*. Grenoble : ELLUG.
- RUWET, Nicolas (1995). Les verbes de sentiment peuvent-ils être agentifs ?, *Langue française*, 105, 28-39.
- SINCLAIR, John McHardy (1991). *Corpus, concordance, collocation*, Oxford: Oxford University Press (Describing English Language).

# ANALYSE BIBLIOMETRIQUE DES RECHERCHES SUR L'ÉCRITURE EN LANGUE SECONDE ET ÉTRANGÈRE EN CHINE<sup>1</sup>

Renlei WANG<sup>a</sup>

[renlei.wang@yahoo.com](mailto:renlei.wang@yahoo.com)

<sup>a</sup>Laboratoire Ligérien de Linguistique, Université d'Orléans, Orléans, France

## 1. Introduction

Par rapport aux recherches sur l'écriture en langue maternelle qui tiennent compte des dimensions affectives, cognitives et psycholinguistiques, Leki, Cumming & Sylva (2008 : 3) rappellent qu'en langue seconde, « early interests in texts and cognitive processes have expanded from simple to more complex perspectives that consider broad-based, social understandings and more inclusive images of L2 writing and writers ». De fait, les éléments socioculturels, la motivation, les représentations et le rapport à l'écrit(ure) au sens large (Barré-de Miniac, 2000 ; Chartrand & Blaser, 2008) jouent un rôle de plus en plus important dans les recherches sur l'écriture en langue maternelle comme en langue seconde et étrangère. Par ailleurs, en faisant les descriptions historiques des recherches sur l'écriture en langue seconde, Matsuda (2003) évoque que celle-ci tend à être considérée comme un sous-domaine des recherches en langue seconde. En tant qu'un champ d'activité indépendant, les recherches sur l'écriture en langue seconde seraient variées d'un pays à un autre. En nous intéressant au sujet de la compétence scripturale des apprenants chinois en langue étrangère, nous nous demandons quelles sont les caractéristiques et quels sont les progrès récents en la matière en Chine.

La bibliométrie, qui utilise des méthodes quantitatives et statistiques pour décrire les modèles de distribution des articles, des champs, des institutions ou des pays à l'égard d'un sujet donné, relève des sciences de l'information ou la bibliothéconomie. Elle est appliquée dans différents domaines, tels que les sciences de gestion, l'ingénierie, l'éducation, etc. ((Li & Chen, 2016 : 8). Dans la présente étude, nous cherchons à nous servir de méthodes de la bibliométrie en ayant pour objectif de révéler de façon quantitative et qualitative les tendances globales des recherches sur l'écriture en langue seconde et étrangère en Chine. Plus précisément, nous ferons une analyse bibliométrique des articles publiés, des auteurs, des revues pour décrire les progrès récents dans ce domaine. De plus, l'analyse par mots-clés sera employée pour révéler les thèmes importants. Nous espérons que cet état des lieux établi pourra guider les chercheurs dans le domaine dynamique des recherches sur l'écriture en langue seconde et étrangère en Chine.

## 2. Constitution du corpus et méthodologie

### 2.1. Constitution du corpus

Les données ont été extraites du système de stockage des données du *Chinese Social Sciences Citation Index* (CSSCI<sup>2</sup>), qui couvre environ 500 revues académiques chinoises en sciences humaines et sociales (SHS). De nos jours, cette base de données est utilisée par de nombreux universités et instituts de recherche en Chine pour l'évaluation des performances et des promotions académiques. D'abord, comme l'anglais, le français, l'allemand, l'italien, l'espagnol, le portugais, le russe, le japonais, le coréen, l'arabe sont les dix langues étrangères les plus populaires en Chine (Western Returned Scholars Association, 2015), nous avons utilisé comme mots-clés "écriture en langue seconde", "écriture en langue étrangère", "écriture

<sup>1</sup> Ce travail est soutenu par *China Scholarship Council*.

<sup>2</sup> Le site internet du CSSCI : <http://cssci.nju.edu.cn>.

en anglais", ou en particulier "écriture en français", "écriture en allemand", "écriture en italien", "écriture en espagnol", "écriture en portugais", "écriture en russe", "écriture en japonais", "écriture en coréen" ou encore "écriture en arabe" pour rechercher, au sein de la base de données du CSSCI, tous les documents de 1998 à 2016<sup>3</sup> qui comprennent ces mots dans la liste des sujets ou des mots-clés. Ensuite, nous avons sélectionné les documents qui sont classés dans la catégorie "article". Enfin, les données en double ont été supprimées et 462 articles liés à l'écriture en langue seconde et étrangère ont été identifiés en vue d'une analyse plus approfondie pour la présente étude.

Processus	Documents trouvés
1. Année : 1998-2016	
2. Critère de sélection : Type de document "article"	
3. Mots-clés : sujet ou mots-clés =	
• "écriture en langue seconde"	70
• "écriture en langue étrangère"	15
• "écriture en anglais"	384 <sup>4</sup>
• "écriture en français"	0
• "écriture en allemand"	0
• "écriture en italien"	0
• "écriture en espagnol"	0
• "écriture en portugais"	0
• "écriture en russe"	4
• "écriture en japonais"	0
• "écriture en coréen"	0
• "écriture en arabe"	0
	Total : 473
4. Critère d'exclusion : Données en double	11 <sup>5</sup>
	Reste : 462

Tableau 3 - Les étapes de sélections

## 2.2. Méthodologie

La méthode bibliométrique que nous adoptons pour cette étude tire profit des technologies modernes en génie informatique, en gestion de base de données et en statistiques. Le terme de bibliométrie est introduit pour la première fois par Pritchard qui le définit comme « the application of mathematics and statistical methods to books and other media of communication » (1969 : 349). On se sert de cette méthode pour une analyse quantitative de la communication scientifique afin de mesurer l'influence académique d'un article ou d'un ouvrage publié. À cet effet, l'analyse « required a multitude of sophisticated techniques including citation analysis, statistical analysis and other quantitative techniques for mapping and measurement of relationships and flows among academic backgrounds, research results, research hotspots or other knowledge-based entities » (Lin, Wu & Hong, 2015: 143). Par ailleurs, la cartographie des connaissances, qui est un domaine relativement récent de la bibliométrie, consiste à mettre en évidence l'évolution et l'état actuel de la recherche par le biais des technologies de visualisation. Sous forme spatiale, elle permet de visualiser les relations entre les disciplines, les articles, les auteurs, etc., en analysant les mots-clés, les citations, leurs cooccurrences, etc., ceci afin d'observer les sujets d'actualité de la recherche, de prévoir l'orientation de développement d'une discipline (Xiao, Qiu, Huang, Li & Feng, 2013).

<sup>3</sup> Le CSSCI enregistre les informations dès 1998 et nous avons recueilli notre corpus en 2017.

<sup>4</sup> Ce serait raisonnable, étant donné le fait que l'anglais est la première langue étrangère en Chine. D'ailleurs, un tiers des Chinois apprend l'anglais.

(cf. <http://edu.163.com/13/1018/18/9BG5DD7H00294KNQ.html>, consulté le 20/05/2019).

<sup>5</sup> Nous avons supprimé les données en double à l'aide du logiciel Citespace.

Dans notre étude, nous nous sommes servis de Citespace pour la cartographie des connaissances. Ce logiciel de visualisation scientifique est développé par le professeur Chen de l'Université de Drexel et appliqué dans l'analyse de cocitation (Chen, 2004, 2006). Il est capable de détecter les mots les plus fréquents dans les articles sur une période donnée. Il est aussi en mesure de traiter non seulement les données anglaises, mais également les données chinoises. Concernant les résultats du traitement de nos données, soit les figures produites, elles sont traduites en français par nous-même.

### 3. Résultats et analyses

#### 3.1. Publication annuelle d'articles

La publication annuelle d'articles se réfère au nombre total d'articles scientifiques qui sont publiés par les revues chinoises en SHS. La figure suivante montre la distribution annuelle des articles au sujet de l'écriture en langue seconde et étrangère recueillis à partir de la base de données du CSSCI. Elle permet de représenter une tendance instable et relativement à la baisse du nombre d'articles. On pourrait diviser cette tendance en trois étapes : au cours de la première étape, soit de 1998 à 2001, on voit une tendance générale à la baisse du nombre d'articles ; de 2001 à 2011, la publication annuelle reconnaît une augmentation et atteint son maximum en 2010 et en 2011, même avec des fluctuations ; dans la dernière étape, le nombre total de publications annuelles baisse depuis 2011, en dépit d'une nette tendance à la hausse en 2015. Comme le CSSCI se met à collecter les données dès 1998, les informations avant cette année sont incomplètes.



Figure 5 - Distribution annuelle d'articles

#### 3.2. Répartition des revues

La diffusion des recherches sur l'écriture en langue seconde et étrangère se fait à travers 93 revues chinoises en SHS. Le tableau 2 présente le Top 10 des titres de revues qui ont publié les articles les plus nombreux en la matière. De 1998 à 2016, sont sortis, dans ces 10 revues, 17 514 articles au total dont 284 concernent l'écriture en langue seconde et étrangère, soit 1,62 % du chiffre total. Les trois premières revues sont *Foreign Language World*, *Computer-assisted Foreign Language Education*, *Foreign Languages and Their Teaching* (respectivement avec 55, 54, 39 articles publiés). Dès lors, on pourrait dire que les recherches sur l'écriture en langue seconde et étrangère sont relativement très peu nombreuses.

De manière évidente, il s'agit principalement des revues portées sur l'enseignement des langues étrangères, l'éducation et la branche d'étude en langues et littérature étrangères. Sauf que *Computer-assisted Foreign Language Education* est consacré aux recherches et à l'enseignement des langues étrangères, en s'appuyant sur les technologies de l'information et de la communication dans tout le système éducatif contemporain, et que *Modern Education Technology* s'applique à étudier l'utilisation de la pédagogie moderne et de la technologie de

l'information moderne pour parvenir à une optimisation de la théorie et de la pratique de l'enseignement, huit autres revues sont animées par les établissements supérieurs, en particulier les universités des langues étrangères. En l'occurrence, c'est en milieu universitaire que l'on s'occupe plus des questions sur les langues étrangères.

Revue	N° de pub	N° total	Pourcentage
1. Foreign Language World	55	924	5,95 %
2. Computer-assisted Foreign Language Education	54	1056	5,11 %
3. Foreign Languages and Their Teaching	39	1869	2,09 %
4. Foreign Language Education	27	5087	0,53 %
5. Modern Foreign Languages	25	878	2,85 %
6. Modern Education Technology	24	3138	0,76 %
7. Foreign Language Research	23	2655	0,87 %
8. Foreign Languages in China	21	783	2,68 %
9. Foreign Language Teaching and Research	16	1124	1,42 %
10. Journal of PLA University of Foreign Languages	11	854	1,29 %
Total	284	17 514	1,62 %

Tableau 4 - Top 10 des titres de revues chinoises en SHS dans le domaine de l'écriture seconde et étrangère en Chine (1998-2016)

### 3.3. Analyse des auteurs

Les chercheurs de pointe sont en mesure de guider le développement de la discipline dans un domaine concerné (Sun, 2012), donc leur dynamisme permet d'y mettre en lumière les recherches en cours ainsi que l'état de développement des recherches.



Figure 6 - Cartographie de la collaboration entre les auteurs

Au total, on compte 372 premiers auteurs. D'une manière visuelle, *Citespace* peut montrer le nombre d'articles et leur date de publication sous forme de cercle annuel ; la circonférence du cercle est proportionnée au volume de publication (Li & Chen, 2016 : 90). D'une part, la figure ci-dessus présente les 20 premiers auteurs qui ont publié deux ou plus de deux articles sur l'écriture en langue seconde et étrangère en Chine. Les six premiers sont en ordre décroissant : Yonglin Yang (9), Na Wang (6), Hongyun Wu (5), Xiaoqing Qin (4), Fang Xu (3) et Yanren Ding (3). D'autre part, la ligne droite qui rejoint les auteurs présente clairement la situation de la collaboration entre ces derniers. Parmi ces 20 premiers auteurs, il n'y a que six équipes de recherche élaborées respectivement par deux chercheurs — entre Na Wang et Hong Zhang, entre Na Wang et Rongjun Liu, entre Fang Xu et Yanren Ding, entre Lifei Wang et Gong Chen, entre Jin Bi et Xiaoqing Qin, entre Hongyun Wu et Runqing Liu (cf. les pointillés dans la figure 2) — et une imposante majorité d'entre eux ont travaillé tout seuls. Autrement dit, la force globale de recherche se manifeste de manière dispersive.

### 3.4. Analyse des tendances de la recherche

Même si les mots-clés ne sont pas nombreux, ils offrent un résumé succinct du thème d'un article, et constituent ainsi un élément important de l'article. Avec *Citespace*, on peut étudier les mots-clés le plus fréquemment indiqués par les articles de manière à aborder différents sujets d'étude. Vu qu'entre 1998 et 2016, le nombre annuel d'articles publiés sur l'écriture en langue seconde et étrangère en Chine est entre 2 et 44, nous avons fait le choix du compromis en retenant les 20 premiers articles les plus cités pour créer un réseau simplifié de mots-clés rattachés à ces articles. Au cours de ce processus, les mots-clés ayant la même signification ont été combinés. Ainsi, en fonction de leur fréquence, il en résulte 57 mots-clés, dont les dix premiers : écriture en anglais (170), écriture en langue seconde (42), enseignement de l'écriture (37), enseignement de l'écriture en anglais (32), enseignement de l'anglais (22), écriture en anglais de niveau universitaire (19), anglais de niveau universitaire (15), corpus (11), écriture scientifique en anglais (9), métacognition (7).

La centralité d'intermédiarité (*betweenness centrality*) est un indice relatif à l'importance des nœuds au sein du réseau de la visualisation. *Citespace* se sert de cet indice pour mesurer l'importance de l'article (ou de l'auteur, de la revue, etc.), et encercle en violet celui dont la centralité est égale ou supérieure à 0,1 (Li & Chen, 2016 : 89). Le tableau 3 présente tous les huit mots dont la centralité est juste égale ou supérieure à 0,1. Ce résultat concorde bien avec celui que nous avons obtenu sur les dix premiers mots-clés les plus fréquents. En effet, les principales recherches sur l'écriture en langue seconde et étrangère en Chine se concentrent sur l'écriture en anglais et l'enseignement de l'écriture en anglais, en particulier en milieu universitaire. Ce phénomène n'est pas difficile à comprendre, parce que les élèves chinois commencent à apprendre l'anglais au cours élémentaire deuxième année de l'école primaire, jusqu'à ce qu'ils finissent leur formation de licence, dès lors, chaque élève reçoit généralement une instruction en anglais au moins pour une période de douze ans.

Mots-clés	Centralité
1. Anglais écrit	0,96
2. Enseignement de l'écriture	0,47
3. Enseignement de l'anglais écrit	0,27
4. Enseignement de l'anglais	0,25
5. Anglais de niveau universitaire	0,20
6. Corpus	0,16
7. Écriture scientifique en anglais	0,15
8. Feedback de l'enseignant	0,13

**Tableau 5 - Top 8 mots-clés dont la centralité est supérieure à 0,1**

Nous avons fait une cartographie de Fuseau horaire (Timezone), qui présente l'évolution des mots-clés dans le cadre des recherches sur l'écriture en langue seconde et étrangère en Chine. Cette cartographie de fuseau horaire permet de mettre en lumière les recherches classiques et récentes ainsi que les liens entre elles. Selon la figure suivante, on peut constater que les recherches se concentrent sur l'écriture en anglais et l'enseignement de l'écriture en anglais avant l'année 2009, et qu'à partir de l'année 2009, de nouveaux sujets, tels que l'analyse du corpus de textes ou de l'écriture académique, ont surgi l'un après l'autre.

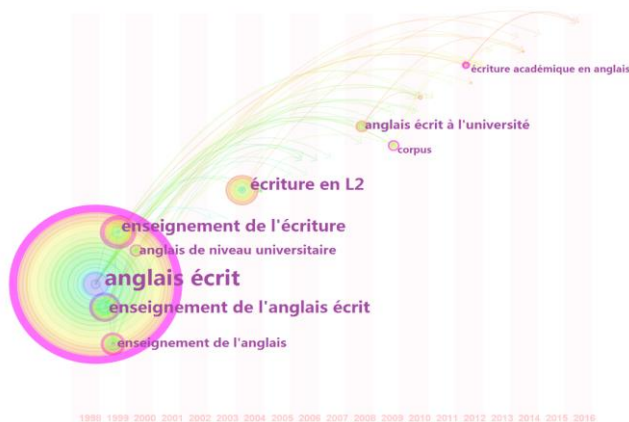


Figure 7 - Cartographie du Fuseau horaire à propos des mots-clés indiqués par les articles<sup>6</sup>

Afin d'obtenir plus d'informations sur l'évolution des mots-clés et leur changement dynamique dans différentes périodes, nous avons divisé la période de référence choisie (1998-2016) en trois cycles de six ou sept ans : 1998-2003, 2004-2009, 2010-2016. Avec *Citespace*, les mots-clés recensés ont ensuite été classés en différentes catégories selon la nature de leur contenu. Enfin, il en ressort trois cartographies du regroupement des mots-clés. Les groupements peuvent inclure des mots-clés communs et nous avons choisi, pour chaque groupement, des mots-clés représentatifs d'un certain sujet — comme le montrent les tableaux 4-6 ci-dessous —, ceci pour faire un résumé des principaux sujets à chaque étape du traitement.



Figure 8 - Cartographie du regroupement des mots-clés<sup>7</sup> (période 1998-2003)

N° de Regroupement	Étiquettes
# 0.	Analyse lexicale/grammaticale/textuelle, linguistique contrastive, influence du chinois, analyse rhétorique, méthode du processus d'écrire
# 1.	Réforme éducative, recherche sur l'enseignement, mode d'enseignement, didactique, enseignement à l'école secondaire
# 2.	Lettre, mémoire, anglais professionnel international, anglais des affaires

Tableau 6 - Description du regroupement (1998-2003)

Durant la première période (cf. figure 4 et tableau 4), les travaux sur l'écriture en langue seconde et étrangère se concentrent sur l'analyse linguistique du texte, la comparaison linguistique, l'enseignement de l'écriture, les genres et l'écriture.

<sup>6</sup> Pour des fins de concision, nous n'avons pas fait apparaître tous les mots-clés sur la cartographie.

<sup>7</sup> Dans la figure 4, le nom de regroupement automatiquement généré est remplacé par les trois points de suspension, car il ne peut pas bien résumer l'ensemble de mots-clés en fonction de leur contenu dans un même regroupement. C'est le même cas pour la figure 5 et la figure 6.



Figure 9 - Cartographie du regroupement des mots-clés (période 2004-2009)

N° de Regroupement	Étiquettes
# 0.	Théorie du postprocessus, métacognition, motivation
# 1.	Compétence de communication, transfert de la langue maternelle, culture, mode de pensée culturelle, rétroaction des enseignants/pairs, évaluation
# 2.	Méthode d'écriture, élaboration du système pédagogique et d'exercices, enseignement assisté par ordinateur, Internet
# 3.	Analyse des erreurs, catégories lexicales, autocorrections, corpus d'écriture

Tableau 7 - Description du regroupement (2004-2009)

Au cours de la deuxième période (cf. figure 5 et tableau 5), on s'est orienté vers le processus rédactionnel, les facteurs qui influencent l'écriture, l'évaluation/rétroaction. Dans cette période, grâce au développement d'Internet, l'ordinateur pour l'enseignement de l'écriture joue un nouveau rôle dans le domaine des recherches sur l'écriture en langue seconde et étrangère en Chine. De plus, l'analyse du texte y occupe toujours une place primordiale.

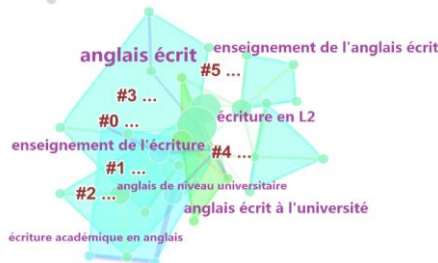


Figure 10 - Cartographie du regroupement des mots-clés (période 2010-2016)

N° de Regroupement	Étiquettes
# 0.	Interaction multimodale, mode d'enseignement, sujet/objet
# 1.	Écriture scientifique, journal, communication par e-mail
# 2.	Compétence lexicale, caractéristique linguistique, différence individuelle, angoisse en face de l'écriture, mémoire de travail
# 3.	Notation automatique, évaluation dynamique, détection des erreurs
# 4.	Feedback des enseignants/pairs, rétroaction en ligne
# 5.	Écriture/bibliothèque numérique, technologie informatique

Tableau 8 - Description du regroupement (2010-2016)

Pendant la dernière période (cf. figure 6 et tableau 6), l'enseignement actif fait jouer un rôle central à l'apprenant en classe. L'analyse du texte fini, du processus d'écriture et de l'évaluation/rétroaction joue encore un rôle important. Par ailleurs, les chercheurs ont prêté une attention particulière à des éléments propres au scripteur, tels que l'angoisse face à l'écriture, la mémoire de travail. Enfin, à l'ère numérique, le développement de la technologie



informatique favorise l'apprentissage-enseignement de l'écriture, d'où la bibliothèque numérique, la plateforme d'écriture en ligne, le système automatique d'évaluation, etc.

De tout ce qui précède, nous pouvons diviser en cinq principales parties les recherches sur l'écriture en langue étrangère et seconde en Chine : recherches sur l'enseignement de l'écriture, recherches sur le produit fini qu'est le texte, recherches sur le processus rédactionnel, recherches sur l'environnement de l'écriture et recherches sur la mémoire de travail.

### **(1) L'enseignement de l'écriture**

Certains chercheurs se sont appuyés sur des théories par rapport à la pratique de l'enseignement de l'écriture, telles que le constructivisme<sup>8</sup> (Guo, 2009b), la métacognition (Wu & Liu, 2004), etc. Certains ont exploré la mise en place de nouvelles didactiques de l'écriture en classe : LSRW (listening-speaking-reading-writing) (Jiang, 2006), approche axée sur le processus d'écriture (Wu & He, 2015), approche de projet, qui s'appuie sur une enquête approfondie à propos d'un sujet du monde réel (Ding & Liu, 2011), etc. D'autres ont montré la place que tient la notion de genre dans l'enseignement de l'écriture, d'où la rédaction de lettres commerciales (Zhao & Zhou, 1999). Avec le développement de la technologie informatique et d'Internet, Guo (2009a) s'est servi du blog pour améliorer l'enseignement de l'écriture en milieu universitaire ; Gui (2007) a analysé l'enseignement de l'écriture dans un environnement à distance à partir de plusieurs aspects, y compris le professeur, le manuel, les ressources, le support technique ; Wang (2014) a fait une étude sur l'enseignement de l'écriture numérique basée sur la plateforme de ressources pédagogiques en rédaction. Ces recherches ont pour objectif un enseignement efficace de l'écriture en langue seconde et étrangère.

### **(2) Le texte fini**

L'analyse du texte fini est principalement constituée de l'analyse des caractéristiques linguistiques et de l'analyse textuelle, y compris l'analyse lexicale ou grammaticale, l'analyse contrastive, l'analyse des erreurs, l'analyse de la rhétorique, etc. Dans ce processus, le problème de genre y joue un rôle important, en particulier dans l'analyse de l'écriture scientifique. Par exemple, Zhang (2015) a discuté des difficultés d'apprenants chinois dans la citation académique ; la recherche de Sun & Wang (2015) vise à examiner la pratique de l'emprunt dans l'écriture académique d'apprenants d'anglais langue étrangère dans le contexte chinois. Ces chercheurs ont fait l'analyse à partir du corpus : soit ils ont choisi leurs données dans de grands corpus existants, soit ils ont rassemblé eux-mêmes des textes pour créer un petit corpus d'analyse.

Les recherches sur l'évaluation du texte regroupent plusieurs directions : les recherches sur les critères de l'évaluation ; les recherches sur l'influence du facteur subjectif sur l'évaluation ; et en particulier les recherches sur l'évaluation automatique du texte. Par exemple, He & Wang (2012) ont construit un système d'évaluation assisté par ordinateur pour les apprenants chinois d'anglais ; Lu, Tan & Peng (2010) ont utilisé à la fois la méthode d'enquête et la méthode quantitative pour analyser l'application du système intelligent de l'évaluation dans l'écriture en anglais de niveau universitaire. En tout cas, le développement du système ou du logiciel d'évaluation a pour but de guider l'enseignement de l'écriture et d'améliorer la compétence scripturale des apprenants en langue étrangère.

Suite à l'évaluation, la rétroaction apparaît comme une nécessité au regard du statut général des textes. Inscrit dans une relation pédagogique qui se construit entre enseignant et étudiant, la rétroaction est l'information donnée soit par l'enseignant à l'étudiant, soit par l'étudiant à l'étudiant. De plus, la comparaison des rétroactions données respectivement par l'enseignant et les pairs a suscité de l'intérêt. Par exemple, Wang & Liu (2012) ont étudié les effets positifs

---

<sup>8</sup> Les études sélectionnées dans les cinq « parties » démontraient davantage nos affirmations.

de la rétroaction des enseignants sur la précision, la maîtrise et la qualité de l'écriture en anglais langue étrangère ; Li & Wei (2012) ont approfondi le rôle que joue la rétroaction donnée par les pairs dans l'écriture en langue seconde, et plus généralement dans l'enseignement de la langue seconde ; Ji (2010) a trouvé que l'efficacité de la rétroaction est proportionnelle à la compétence scripturale. Dans le premier cas, l'efficacité de la rétroaction des enseignants varie dans le même sens : plus la compétence scripturale est élevée et plus l'efficacité de la rétroaction des enseignants augmente. Dans le second cas, l'efficacité de la rétroaction des pairs varie dans un sens opposé : moins il y a de compétence scripturale, plus grande est l'efficacité de la rétroaction des pairs.

### **(3) Le processus rédactionnel**

Depuis 1980, les recherches, qui sont centrées sur le texte fini, sont remises en question par la naissance d'une approche focalisée sur les opérations effectuées par le sujet écrivant, d'où la prise en compte du processus d'écriture (Flower & Hayes, 1980).

Les recherches sur le processus de rédaction portent sur différentes étapes dans un tel processus, sur les processus cognitifs des apprenants, sur le rôle qu'y joue la langue première, etc. Par exemple, Zhang (2006) a cherché à discuter du processus implicite dans lequel se noue une série de relations complexes entre le sujet scripteur, son texte, et le sujet lecteur ; avec la méthode de la pensée à voix haute, Xiu & Xiao (2004) se sont engagés à découvrir la relation entre les processus cognitifs d'écriture et la note chiffrée pour l'évaluation donnée par le professeur ; la recherche de Chen & Wang (2010) vise à étudier 16 apprenants scripteurs dont la spécialité est l'anglais, pour quantifier l'utilisation de la langue première dans l'écriture en langue seconde et explorer ses variations avec les activités de réflexion et la maîtrise de la langue seconde. Bref, les recherches sur le processus rédactionnel en langue seconde et étrangère sont peu nombreuses par rapport à celles sur l'analyse du texte fini.

### **(4) L'environnement de l'écriture**

Les recherches sur l'environnement de l'écriture se rapportent aux facteurs internes et externes qui ont un effet sur l'écriture des apprenants et sur leur compétence scripturale, y compris l'angoisse, la culture, la différence de pensée, etc. Par exemple, Wang & Fei (2008) ont confirmé que la langue et la culture sont intimement liées l'une à l'autre et que l'écriture en langue étrangère ne peut être enseignée sans culture ; la prise en compte de la différence culturelle contribue à réduire les conflits culturels dans les activités d'écriture en langue étrangère. Zhou & Tang (2010) ont montré que l'angoisse est négativement corrélée avec la note de composition, mais positivement corrélée avec l'interférence avec la langue maternelle.

Écrire en langue seconde et étrangère est un processus complexe : en plus du mode de pensée en langue première, il existe encore différents facteurs qui interviennent dans un tel processus, tels que la compétence scripturale du scripteur, le niveau de la maîtrise d'une langue étrangère, la consigne de l'écriture, etc.

### **(5) La mémoire de travail**

Il s'agit d'un nouveau domaine dans les recherches sur l'écriture en langue seconde et étrangère en Chine. Yan (2010) a examiné les différences individuelles dans les stratégies de révision en matière d'écriture en langue seconde et comment elles sont en relation avec la mémoire de travail. Li (2015) a expliqué et analysé les relations et les interactions entre la compétence, la mémoire de travail, la motivation et différentes étapes de l'écriture en langue seconde. En tout cas, ces recherches portent sur les différences individuelles dans l'activité d'écriture.

## 4. Conclusion

Dans cet article, après avoir mené une analyse bibliométrique sur les recherches sur l'écriture en langue seconde et étrangère en Chine entre 1998 et 2016 en nous basant sur 462 documents extraits de la base de données CSSCI, nous en tirons les conclusions suivantes.

En premier lieu, les recherches liées à l'écriture ne sont pas nombreuses par rapport à d'autres recherches en contexte de langue seconde et étrangère en Chine. En second lieu, il existe une collaboration insuffisante entre les chercheurs. En dernier lieu, les principaux sujets d'étude sont l'enseignement de l'écriture, l'analyse et l'évaluation du texte fini, le processus rédactionnel, l'environnement de l'écriture et la mémoire de travail. Cependant, il y a encore des problèmes, tels qu'un accent trop mis sur les résultats de l'écriture, d'où une négligence du processus de l'écriture ; dans l'activité d'écriture, on n'a pas pris en compte le rapport à l'écrit (Barré-de Miniac, 2000 ; Chartrand & Blaser, 2008) du sujet écrivant ou de l'enseignant, ainsi que la réécriture mise en jeu dans le processus scriptural grâce aux apports de la génétique textuelle (Lafont-terrano, 2009).

Nous espérons qu'en donnant une image générale des recherches sur l'écriture en langue seconde et étrangère en Chine et identifiant les tendances les plus importantes et influençables qui se sont produites dans ce domaine, ce travail servira de sources d'inspiration pour d'autres chercheurs en la matière.

Comme les résultats de cette étude reposent sur l'analyse objective des données, elle est peu influencée par l'expérience subjective de l'auteur et rend ces résultats stables et objectifs. Cependant, en raison des autorisations requises pour la base de données et des limitations fonctionnelles de *Citespace*, une seule utilisation du CSSCI comme source de données pourrait conduire à des descriptions partielles de toute situation réelle des recherches sur l'écriture en langue seconde et étrangère en Chine, ce qui pourrait avoir un impact sur la pertinence de la présente étude.

### Références bibliographiques

- BARRE-DE MINIAC, Christine (2000). *Le rapport à l'écriture : aspects théoriques et didactiques*. Villeneuve-d'Ascq : Presses universitaires du Septentrion.
- CHARTRAND, Suzanne-G. & BLASER, Christiane (2008). Du rapport à l'écriture au concept didactique de capacités langagières : apports et limites de la notion de rapport à l'écrit, in *Le rapport à l'écrit. Un outil pour enseigner de l'école à l'université*, Chartrand, Suzanne-G. & Blaser, Christiane (Eds). Namur : Presses universitaires de Namur, 107-127.
- CHEN, Chao-mei (2004). Searching for intellectual turning points: Progressive knowledge domain visualization, *Proceedings of the National Academy of Sciences*, 101(1): 5303-5310.
- CHEN, Chao-mei (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature, *Journal of the American Society for Information Science and Technology*, 57(3): 359-377.
- CHEN, Xiao-xiang & WANG, Yang (2010). A Quantitative Analysis of L1 Use in L2 Writing Process, *Journal of Hunan University*, 24(6): 89-93.
- DING, Yan-ren & LIU, Hai-ping (2011). The project method in English writing instruction, *Foreign Language World*, 2: 11-13.
- FLOWER, Linda S. & HAYES, John R. (1980). The dynamics of composing: Making plans and juggling constraints, in *Cognitive Processes in Writing*, GREGG Lee W & STEINBERG Erwin R (Eds). Hillsdale, N.J.: L. Erlbaum Associates, 3-20.
- GUI, Ying (2007). On Teaching of English Writing in Distance Environment: A Case Study of BFSU Online College, *Modern Distance Education*, 6: 19-22.
- GUO, Xiao-ying (2009a). Use of blog to improve teaching of college English writing, *Modern Foreign Languages*, 32(3): 314-321.

- GUO, Xiao-ying (2009b). A college English Writing Teaching Experiment Based on Constructivism Theory, *Foreign Languages in China*, 6(4): 72-76.
- HE, Jun-jie & WANG, Fang (2012). Automated Grammar Analysis and Computer-assisted Essay Scoring, *Foreign Languages and their teaching*, 6: 61-65.
- JIANG, Xian-ju (2006). LSRW: A Mode of L2 Writing Teaching, *Foreign Language Teaching Abroad*, 3: 50-53, 30.
- JI, Xiao-ling (2010). A Comparison of Peer Review and Teacher Review in the Writing Course for English Majors, *Journal of PLA University of Foreign Languages*, 33(5): 60-65.
- LAFONT-TERRANOVA, Jacqueline (2009). *Se construire, à l'école, comme sujet-écrivain : l'apport des ateliers d'écriture*. Namur, Belgique : Presses universitaires de Namur.
- LEKI, Ilona, CUMMING, Alister & SYLVA, Tony (2008). *A Synthesis of Research on Second Language Writing in English*. New York : Routledge.
- LI, Jie & CHEN, Chao-mei (2016). *Citespace: Text Mining and Visualization in Scientific Literature*. Beijing : Capital University of Economics and Business Press.
- LI, Shao-peng (2015). On the Research of Individual Differences in the L2 Writing Development, *Language Teaching and Linguistic Studies*, 6: 27-34.
- LI, Xu-Kui & WEI, Xin-feng (2012). Research on the Effect of Peer Feedback on L2 Writing: A Review and Its Implications, *Journal of Ocean University of China*, 3: 104-109.
- LIN, Zhuo, WU, Cheng-zhen & HONG, Wei (2015). Visualization analysis of ecological assets/values research by knowledge mapping, *Acta Ecologica Sinica*, 35(5): 142-154.
- LU, Yan-hui, TAN, Fu-min & PENG, Shun (2010). Intelligent Writing System in College English Writing: An Empirical Study, *Modern Education Technology*, 20(6): 56-58.
- MATSUDA, Paul Kei. (2003). Second language writing in the twentieth century: A situated historical perspective, in *Exploring the Dynamics of Second Language Writing*, KROLL Barbara (Eds). Cambridge : Cambridge University Press, 15-34.
- PRITCHARD, Alan (1969). Statistical Bibliography or Bibliometrics, *Journal of Documentation*, 25(4): 348-349.
- SUN, Hai-sheng (2012). Author Keyword Co-occurrence Network Analysis: An Empirical Research, *Journal of Intelligence*, 9: 63-67.
- SUN, Yan-shu & WANG, Jun-ju (2015). A Study on Textual Borrowing in English Academic Writing, *Foreign Languages and their teaching*, 6: 36-41, 48.
- WANG, Gui-zhi & FEI, Qian (2008). The Impact of Cultural Codes on Interlanguage Writing, *Journal of Northeast Normal University*, 5: 139-143.
- WANG, Na (2014). An exploratory of Digital Teaching Based on Experiencing English – Writing Teaching Resource Program, *Modern Education Technology*, 4: 52-59.
- WANG, Ying & LIU, Zhen-qian (2012). A Study of the Effects of Teacher Feedback on Accuracy, Fluency, Complexity and Quality of EFL Writing, *Foreign Language Education*, 33(6): 49-53.
- Western Returned Scholars Association. (2015). The Employment Prospects for Nine Foreign Languages, *Overseas Students*, 11: 20-21.
- WU, Rong-hui & HE, Gao-da (2015). A Probe into Teaching Model of ESL Writing in Process Approach with IT, *Modern Education Technology*, 25(4): 65-70.
- WU, Hong-yun & LIU, Run-qing (2004). A metacognitive framework for L2 writing: A factor analysis approach, *Foreign Language Teaching and Research*, 36(4): 187-195.
- XIAO, Ming, QIU, Xiao-hua, HUANG, Jie, LI, Guo-jun & FENG, Zhao-hui (2013). Comparison of software tools for mapping knowledge domain, *Library Journal*, 3: 61-69.
- XIU, Xu-dong & XIAO, De-fa (2004). A think-aloud probe into the relationship between cognitive processes and achievement s in TEM-8 writing, *Foreign Language Teaching and Research*, 36(6): 462-466.
- YAN, Rong (2010). Revising Strategies in SLW: Individual Differences and Their Relations with Working Memory, *Foreign Language Education*, 31(6): 55-59.

ZHANG, Li-yin (2015). A Study of Chinese Learners' Difficulties in Academic Citation: Based on an Analysis of Relevant Results Home and Abroad in Recent Ten Years, *Foreign Languages and their teaching*, 5: 22-28.

ZHANG, Sheng-lin. (2006). The Implicit Process in the Process Writing, *Foreign Languages and their teaching*, 4: 28-30.

ZHAO, Chong-hua & ZHOU, Rui-qi (1999). Writing Business Letters of Complains, Claims and Settlements, *International Economics and Trade Research*, 5: 64-68.

ZHOU, Bao-guo & TANG, Jun-jun (2010). An Empirical Study of the Influence of L2 Writing Apprehension on L2 Writing Process, *Foreign Language Education*, 31(1): 64-68.

## **LE TRAITEMENT DE TRANSCRIPTIONS D'ÉCRITS SCOLAIRES : UN PROTOCOLE D'ANNOTATIONS POUR ANALYSER LES INTERVENTIONS DES ENSEIGNANTS**

Sara MAZZIOTTI<sup>a</sup>

[sara.mazziotti@studio.unibo.it](mailto:sara.mazziotti@studio.unibo.it)

<sup>a</sup>Laboratoire CLESTHIA, Paris, France

À l'heure de la digitalisation, la linguistique bénéficie autant que d'autres disciplines d'innovations de pointe qui facilitent et améliorent les études de corpus. Cependant, l'apport humain demeure essentiel notamment dans le cas de corpus d'écrits scolaires et se concrétise lors de l'étape de transcription, nécessaire pour en étudier ses spécificités. Dans notre corpus d'écrits d'élèves de CE2 et de CM2 italiens et français, l'application de balises qui identifient toutes les procédures d'écriture et les interventions et commentaires de l'enseignant, nous permet de passer du manuscrit au fichier exploitable par des logiciels de textométrie. Ainsi, il est possible de dessiner des profils d'« enseignants-correcteurs » grâce à l'extraction et le calcul automatique de toutes les interventions de l'enseignant. Le linguiste Jean-Luc Pilorgé (2010) avait distingué cinq « postures de correction » en analysant justement le type d'intervention de l'enseignant et l'aspect du texte de l'élève sur lequel elle porte. Nous nous baserons sur cette distinction pour explorer les pratiques de correction des enseignants italiens et français et pour en évaluer successivement la prise en compte de la part de l'élève lors d'une consigne de réécriture.

### **1. Constitution d'un corpus d'écrits scolaires français et italiens**

À partir d'un corpus de productions écrites recueillies dans deux écoles italiennes et deux écoles françaises, nous nous sommes intéressés aux pratiques de correction des enseignants français et italiens en CE2 et en CM2 et à l'incidence des deux systèmes linguistiques sur la fréquence de certaines postures de correction. Pour cette étude expérimentale, nous avons soumis aux élèves participant au projet de recherche une consigne de rédaction commune, puis nous avons demandé aux enseignants de corriger les productions écrites selon leurs propres habitudes correctives. Ensuite, nous avons donné aux élèves la possibilité de relire leurs textes corrigés, quelques jours après l'avoir écrit ; et enfin d'en rédiger une version définitive. La consigne était « Que feras-tu quand tu seras grand ? Raconte une de tes journées » ; le temps à disposition pour la rédaction du brouillon et ensuite de la version définitive était d'une heure pour chacune des deux séances. Des métadonnées concernant la profession des parents et la langue maternelle des élèves, ainsi qu'un questionnaire écrit destiné aux enseignants ont été également recueillis. L'enseignant a toujours été présent dans la classe lors du recueil et il était autorisé à répondre aux doutes des élèves et/ou à les aider en cas de blocage. Les élèves avaient également accès aux dictionnaires et aux affiches dans la classe qui rappelaient des règles de conjugaison ou d'accord verbal. En résumé, nous disposons actuellement du brouillon et de la version définitive d'environ 250 élèves, en sachant que le recueil s'achèvera en décembre 2019.

Successivement, chaque version du texte a été transcrite et annotée selon le protocole de transcription utilisé dans le cadre du projet ECRISCOL (Doquet) qui se donne pour objectif de constituer une base de données ouverte en ligne d'un corpus scolaire et d'associer à chaque manuscrit sa transcription génétique présentée en version TEI et les métadonnées concernant l'école, le devoir et l'élève-scripteur. En tenant compte de l'aspect génétique du texte, de l'aspect temporel de chaque procédure d'écriture et d'une duplicité de voix, notamment celle

de l'élève et celle de l'enseignant-correcteur, les corpus scolaires présentent des spécificités que la transcription et l'annotation peuvent mettre en évidence. Bien que chacune de ces spécificités peut être observable par quiconque sur les copies constitutives de notre corpus, l'intérêt de l'outil de transcription réside dans la possibilité de faire ressortir et de mettre en exergue des informations telle que la construction du processus d'écriture de l'élève, l'organisation de son flux d'idées, l'impact des corrections apportées par l'enseignants chez l'élève et la diversité de ces corrections portant sur différents aspects de la langue et du texte. De plus, une transcription difficile, riche en enchâssements de procédures d'écriture et dans laquelle le taux d'intervention de l'enseignant est très élevé, peut suggérer un problème de lisibilité et donc un obstacle pour l'élève dans la phase de réécriture. Souvent, l'annotation même de ce type de manuscrit devient presque illisible en confirmant la présence excessive du stylo de l'enseignant : seule la conversion en version TEI, destinée à la lecture et à l'exploitation par les usagers, rend lisible ce type d'annotations, comme dans le cadre du projet ECRISCOL.

Dans notre étude, la transcription et l'analyse des productions écrites seront destinées en particulier à l'identification des profils de chaque enseignant-correcteur, afin de les comparer avec leurs témoignages et considérations personnelles et d'évaluer l'impact de chacune de ces postures de correction chez les élèves. N'étant qu'au début de l'analyse de notre corpus, nous considérerons ici, d'une part, l'importance accordée à la transcription et à l'élaboration d'un protocole très détaillé pour traiter un corpus scolaire, et, d'autre part, quelques exemples d'exploitation automatique des transcriptions.

## **2. Transcription et annotation des manuscrits : deux vrais outils d'analyse**

Afin d'accéder au contenu des productions écrites des élèves et l'analyser automatiquement à l'aide des outils numériques de textométrie, il est nécessaire de transcrire le texte selon un protocole de transcription très précis. Plusieurs groupes de recherche en France qui s'intéressent aux écrits scolaires ont des exigences de transcription similaires et se confrontent avec les mêmes problèmes liés à la transcription des copies. Dans le cadre du projet E-CALM, financé par l'ANR et coordonné par Claire Doquet, quatre laboratoires (CIRCEFT-ESCOL, LIDILEM, CLESTHIA et CLLE-ERSS) sont en train de collaborer pour constituer, mettre à disposition de la communauté scientifique et étudier un vaste corpus d'écrits d'élèves et d'étudiants. L'idée fondamentale de cette coersion entre les différents laboratoires participant au projet est de rendre les transcriptions et les annotations les plus homogènes possible, afin que chaque laboratoire ait le même protocole, et ce, dans l'optique d'explorer ce corpus « dans des analyses sociologiquement contextualisées » et « d'étudier les modalités de l'écriture dans les avant-textes ». Notre corpus qui intégrera les données déjà collectées et disponibles en ligne dans le cadre du projet ECRISCOL respecte les objectifs suivants :

1. les quatre écoles participant au recueil proviennent de contextes sociologiques comparables ;
2. les productions écrites finales des élèves sont étudiées en comparaison avec le premier texte produit et corrigé par l'enseignant.

Mais quelles spécificités présente un corpus scolaire ?

Tout d'abord, une copie d'élève conserve les traces de la construction du processus d'écriture : tout ajout, rature, remplacement, déplacement et dessin est signalé dans nos transcriptions ; mais aussi elle conserve les traces de la co-construction de l'écriture, en particulier dans le cas des versions finales qui ont été précédées d'un brouillon sur lequel l'enseignant a apporté des corrections ou des commentaires. Il est donc nécessaire de signaler, d'une part, la multiplicité

des voix qui interviennent au sein de la copie et, d'autre part, la temporalité derrière l'action de ces acteurs.

Il ne serait pas inutile de préciser qu'il existe une distinction au sein de notre projet de recherche entre les fichiers texte appelés « transcriptions » et les fichiers texte appelés « annotations ». Le traitement automatique est mis en place à partir des annotations qui sont des transcriptions des manuscrits incluant une correction de toute forme non normée : si dans un premier temps, nous transcrivons le texte de l'élève tel qu'il apparaît dans le manuscrit ; dans un deuxième temps, nous y ajoutons la correction des erreurs d'orthographe et nous effectuons des modifications de certaines structures du protocole de référence, en préparation de l'exploitation à l'aide des outils de textométrie. Regardons l'extrait d'un manuscrit de CM2 (figure 1) et l'alignement de sa transcription et de son annotation :

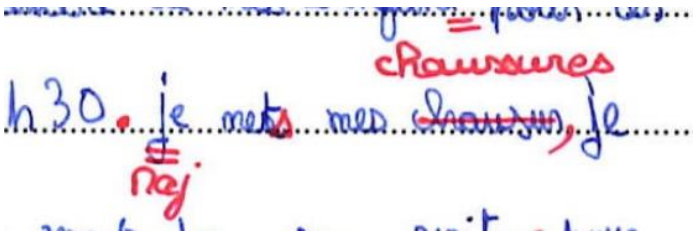


Figure 1 - Extrait d'un manuscrit français de CM2 corrigé.

Transcription	Annotation
{T2P#j}{T2P#maj}e me[s]<t><T2P#s> mes	<{T2P#j}{T2P#maj}e><Je> me@[s]/t//@//T2P#s//
{T2P#chausur}<T2p#chaussures> <T2P#,> je£	mes @[T2P#chausur//T2P#chaussures//@ //T2P#// je
	£

Tableau 1 - Comparaison entre le segment transcrit et le segment annoté de l'extrait de copie en figure 1

Les chevrons (< >), qui signalent dans la transcription les ajouts et donc aussi l'ajout à l'intérieur d'un remplacement, effectué ici d'abord par l'élève (remplacement du « s » par le « t ») et ensuite par l'enseignant (remplacement de « chausur » par « chaussures »), deviennent des slashes dans l'annotation. Dans cette dernière, nous délimitons la totalité de cette procédure avec @, symbole qui indique le début et la fin du remplacement. De plus, toutes les erreurs d'orthographe et de majuscules (ici, nous avons « Je ») sont corrigées à l'aide des chevrons, toujours dans la phase d'annotation (<je><Je>). Le protocole de transcription des soulignements et des commentaires de l'enseignant, que nous expliquerons plus loin, ne varie pas lors du passage aux annotations : dans ce cas, nous situons le soulignement du « j » minuscule et du commentaire « maj » (« majuscule ») à l'intérieur du premier segment entre chevrons que nous avons corrigé avec « Je ». Le premier segment entre chevrons indique en effet le mot écrit par l'élève et le deuxième sa forme normée ajoutée par le transcripteur. La phase de passage des transcriptions aux annotations nous permettra de signaler et de corriger ainsi toute forme non normée dans les transcriptions et d'aboutir à un fichier qui sera converti en format TEI.

Dans la figure 2, un extrait d'une copie de niveau CE2 faisant partie de notre corpus italien. Ici, nous retrouvons en noir le premier jet de l'élève et en rouge les corrections de l'enseignant situées dans le texte, en marge et en bas de page.



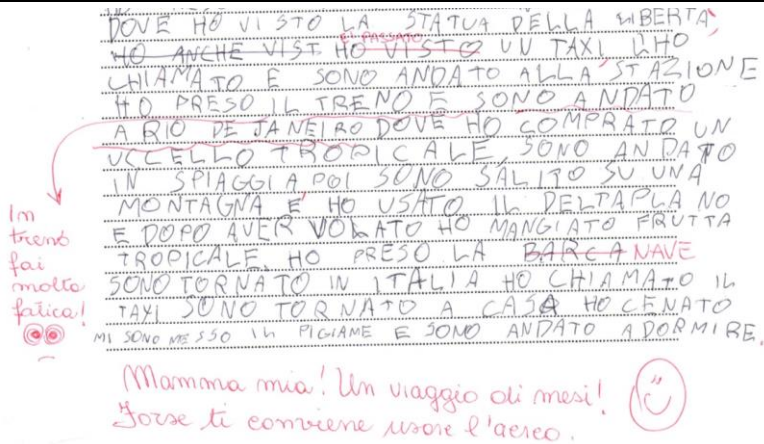


Figure 2 - Extrait d'un manuscrit de CE2 italien corrigé.

Dans ce cas, les interventions de l'enseignant sont très hétérogènes : on retrouve des corrections orthographiques (ajout de l'accent sur le « a » final de « libertà »), des soulignements, des remplacements lexicaux (ex : [barca]<nave>), des commentaires sur le contenu (en marge, pour souligner qu'il est difficile d'aller à Rio De Janeiro en train), des appréciations globales (en bas de page, et qui insistent sur l'idée qu'il s'agit d'un long voyage et qu'il faudrait peut-être prendre l'avion) et des dessins (des smileys). En ce sens, la transcription n'est pas seulement un outil de constitution des données en vue du traitement automatique des textes, elle est aussi un moyen de préparation à la réflexion, un instrument pour rentrer dans le texte et adopter à la fois le point de vue de l'élève et celui de l'enseignant. En effet, nous avançons déjà des hypothèses concernant la lisibilité du texte après la correction de l'enseignant et nous remarquons les spécificités des commentaires de cet enseignant qui suggère implicitement à l'élève de faire attention à sa représentation du monde plus qu'à la forme. Il pourrait s'agir de la posture du « lecteur-naïf » définie par J-L. Pilorgé qui « commente l'action inventée par l'élève en référence au monde d'aujourd'hui » (Pilorgé, 2010). Cependant, nous avons besoin, pour cela, d'abord d'une analyse de toutes les interventions de l'enseignant sur la copie : c'est pour cela qu'il est important de distinguer dans notre protocole tout soulignement, commentaire et remaniement direct par l'enseignant sur le discours de l'élève. Ainsi, nous aboutirons à des annotations qui sépareront ces trois types d'interventions et qui pourront être repérées automatiquement.

### 3. Des annotations au logiciel de textométrie

#### 3.1. Les différents types d'interventions des enseignants à l'écrit

À partir du manuscrit (figure 2), nous allons proposer l'annotation et ensuite les trois types d'interventions de l'enseignant que nous pourrions rechercher, compter et classer automatiquement :

*dove ho visto la statua della liberta*@[T2P#a]//[T2P#à,]@[£

[ho anche vist] @[T2P#ho visto]//[T2P#è passato]@[ un taxi //T2P#,,/ l'ho£

chiamato e sono andato alla stazione£

ho preso {T2P#il treno e sono andato£

a Rio de Janeiro} [T2P#In treno fai molta fatica!]{T2P#DESSIN} dove ho comprato un£

uccello tropicale, sono andato£

in spiaggia //T2P#,,/ poi sono salito su una£

montagna e ho usato il deltaplano£

*e dopo aver volato ho mangiato frutta*

*tropicale. Ho preso la* [T2P#barca]//T2P#nave//£

*sono tornato in Italia ho chiamato il*

*taxi sono tornato a casa*[o]//a//£ ho cenato

*mi sono messo il <pigiamе>\_<pigiamа> e sono andato a dormire* //T2P#//§

*{T2P#Mamma mia! Un viaggio di mesi! Forse ti conviene usare l'aereo.}{T2P#DESSIN}§*

- Ajout de l'accent sur le « a » final de « libertà » (ligne 1). Il s'agit d'un cas de remplacement du « a » sans accent avec le « a » accentué qui est annoté comme une suppression suivie d'un ajout et limité par ce symbole marquant le remplacement ®. Dans notre protocole de transcription, les interventions de l'enseignant sont signalées avec un « P » précédé de l'indication temporelle « T2 » (temps 2, par rapport au premier jet de l'élève) :

libert®[T2P#a]//T2P#a//®

- Soulignement d'une phrase (ligne 4 et 5). La phrase est limitée par des accolades et précédée toujours par la marque temporelle T2, l'indication de l'acteur P et une lettre grecque, le sigma :

{T2Pσ#ho preso il treno e sono andato a Rio de Janeiro}

- Commentaire en marge ou en bas de page. Tout ajout ou commentaire en marge se doit d'être restitué dans le texte, donc nous allons l'insérer juste après la phrase soulignée entre accolades, mais sans le petit sigma :

{T2P#In treno fai molta fatica!}

Pour conclure, avec cette dernière structure nous transcrivons aussi la présence d'un smiley, toujours dans l'extrait du manuscrit proposé dans la figure 1, (ou si présent aussi un tableau, un schéma ou tout autre dessin) qui dans ce cas apparaît juste après le commentaire, toujours en marge : {T2P#DESSIN}

Comme nous l'avons vu plus haut, l'extraction automatique de toutes les procédures d'écritures dans le brouillon et dans la version finale et les interventions des enseignants précédemment annotées dans nos transcriptions nous est utile lors du classement des différents profils de correction qui toutefois n'ont pas l'ambition d'associer un seul type d'intervention à chaque enseignant. Puisqu'au sein d'une même copie nous retrouvons plusieurs types de correction, il est nécessaire de les extraire automatiquement pour les calculer et les analyser. Ceci est possible grâce à un script que nous avons créé, mais que nous ne présenterons pas ici ou à l'aide du logiciel d'analyse textométrique *LeTrameur* conçu par Serge Fleury qui nous montre également le contexte gauche et droit. Cette dernière démarche se révèle très utile également pour investiguer le discours de l'élève.

En effet, le logiciel nous permet de sélectionner un mot-clé (prenons l'exemple d'un mot utilisé dans la consigne, comme « grande ») et de comparer la construction du texte chez les filles et chez les garçons, ou bien de la comparer chez les élèves français et chez les élèves italiens. Voici un exemple d'interface du concordancier de la version en ligne *iTrameur*, à partir du mot « grande » dans l'une des quatre classes italiennes de niveau CE2 (figure 3).

**LeTrameur** Analyse textométrique de données

Afficher  items Recherche :

N°	Partie	Contexte Gauche	Pôle	Contexte Droit
1	EC-CE2-2017-FOR1-D1-E10-V1	a casa non (d)n(/d) stanca per la cena . ] § § § (d)Io da grande, vorrei esse(/d) § § Io da	grande	(i)T2#,(i) vorrei essere una pasticceria . E nel mio negozio , vorrei come
2	EC-CE2-2017-FOR1-D1-E11-V1	di più , ma di cosa farai in una giornata quando sarai	grande	} § / T3#Io da grande sarò l ' archeologa . § Oggi dovrò andare a scavare
3	EC-CE2-2017-FOR1-D1-E12-V1		sarò grande	farò la (an)(b)pittrice(/b)(a)pittrice(/a)(/an) perché mi piace (an)(b)disegn(i)T2#n(i)are(/b)(a)disegnare(/a)(/an) , (d)=XXX#(/d) andrò { T2P#ha } (i)T2#a(/i)
4	EC-CE2-2017-FOR1-D1-E14-V1	racconto una giornata di quando sarò	grande	. § Il 12 luglio , il mio compleanno . § Io forse farò la scrittrice

**Figure 3 - Capture d'écran de la concordance du mot « grande » dans la version en ligne du logiciel *LeTrameur***

Nous repérons facilement, dans le contexte droit, les premières idées que la consigne a suscitées chez l'élève, comment il les a organisées pour débiter son texte et quels temps verbaux il leur a associés.

### 3.2. *Le passage du brouillon à la V2 : la prise en compte des corrections*

L'étude des différents types d'intervention des enseignants ne se limite pas aux textes sur lesquels ces corrections ont été apportées, mais elle s'étend également à l'analyse de la version finale produite par l'élève. L'hypothèse de base de cette recherche considère les enseignants français comme étant davantage focalisés sur le langage, sur l'aspect orthographique, sur la forme qu'ils protègent et corrigent comme des « gardiens du code » (Pilorgé, 2010). En revanche, les enseignants italiens aborderaient plus rapidement dans la scolarité des élèves des aspects tels que la cohérence et corrigeraient plus souvent la construction du texte ou la vraisemblance du contenu. Nous avançons l'hypothèse que cela est dû initialement aux caractéristiques intrinsèques aux deux systèmes linguistiques : une plus forte correspondance graphie-phonie propre à la langue italienne accélère l'apprentissage de l'orthographe et permet aux enseignants de fournir des conseils finalisés à l'amélioration globale du texte initial et non seulement de la forme.

Pour vérifier cette hypothèse et surtout pour évaluer quelle est la posture de correction la plus propice dans le cadre d'un exercice de réécriture de ce type, nous allons comparer la transcription du brouillon corrigé et la transcription de la version définitive, à l'aide du logiciel *mkAlign* conçu par Serge Fleury. Il s'agit d'une première étape d'analyse des deux versions pour évaluer le taux de remaniement du brouillon : le logiciel crée un tableau qui signale en rouge tous les éléments supprimés lors du passage à la version finale et en vert tous les éléments que l'élève a ajoutés. Avec cette mise en comparaison des deux versions, il est possible de déterminer la prise en compte des corrections de la part de l'élève, en sachant que la consigne de réécriture leur donnait aussi la possibilité de ne pas réécrire le premier texte corrigé, mais d'en rédiger un nouveau s'ils le préféraient. Les élèves choisissent-ils de ne pas retravailler le brouillon et de proposer un nouveau texte quand le premier est surchargé de corrections et/ou assez illisible ? Essayent-ils, en revanche, de prendre en compte les corrections et les commentaires de l'enseignant quand ils portent plutôt sur la globalité du texte ?

#### 4. Quelques résultats

Comme précédemment souligné, la transcription et ensuite l'annotation du corpus scolaire constituent une première phase importante de l'analyse de nos données. Nous en déduisons certains éléments qui pourraient être utiles au moment de l'exploitation automatique des annotations qui intervient dans un deuxième temps. Nous ne sommes pas encore en mesure de proposer des chiffres exacts concernant les pratiques de correction des enseignants dans notre corpus, mais nous remarquons cependant une insistance des enseignants français sur l'aspect orthographique et grammatical de la langue ainsi qu'une surcharge de signes dans la façon de corriger les productions écrites de leurs élèves. En outre, les corrections orthographiques sont moins fréquentes en contexte italien car, à parité de niveau, le système linguistique fait en sorte qu'il y ait moins d'erreurs d'orthographe. Nous avons choisi de ne proposer ici que deux extraits de copies corrigées, l'une italienne (figure 1) et l'autre française (figure 4), dans le but de montrer l'hétérogénéité des pratiques de correction. La figure 4 présente un extrait de texte, écrit par un élève de CE2 d'une des deux écoles françaises participant au projet, qui met en évidence la multiplicité de traits, de signes et de couleurs au sein d'une même copie et qui devraient donner des pistes à l'élève en vue de la réécriture du brouillon.

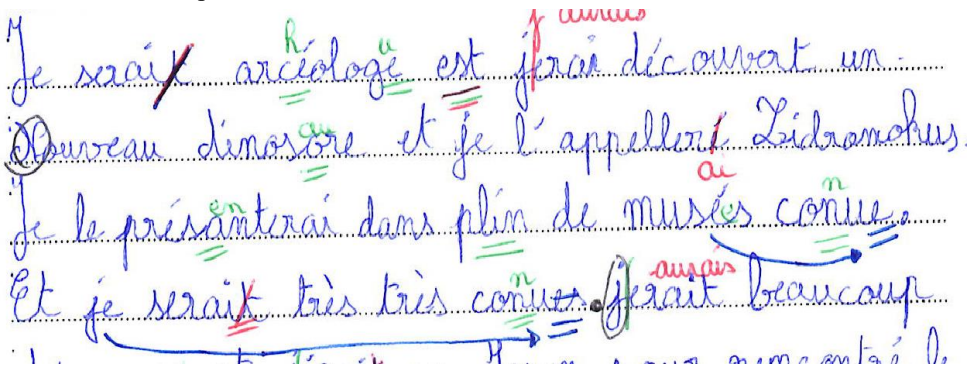


Figure 4 - extrait d'un manuscrit d'un élève de CE2 corrigé.

Si nous comparons ce manuscrit au manuscrit italien de la figure 1, nous remarquons une intervention massive de la part de l'enseignant qui se sert de différentes couleurs pour suggérer à l'élève sur quel aspect porte l'erreur (lexical, grammatical, manque de majuscule). Nous signalons que dans une étude précédente à cette recherche menée dans deux classes françaises de CM2, nous avons constaté qu'une surcharge d'interventions de l'enseignant inhibait le travail de réécriture de l'élève, lequel procédait à une mise au propre du brouillon corrigé en essayant de recopier toutes les corrections du maître. Même si une telle fréquence d'interventions n'a pas été retrouvée dans notre corpus italien, nous avons repéré toutefois la posture de « gardien de code » chez les enseignants italiens également et dans tous les deux niveaux scolaires analysés, avec des nuances différentes que nous allons rapidement présenter maintenant.

Premièrement, les quelques cas de remplacements lexicaux que l'on a retrouvés dans le manuscrit italien (figure 1) sont en réalité très fréquents et imputables à une volonté des enseignants de faire varier le plus possible le vocabulaire dans le texte de l'élève. Il ne s'agit pas uniquement de fournir des synonymes pour éviter les répétitions : pourrions-nous avancer l'hypothèse d'une véritable tendance que les enseignants italiens font émerger dans leur pratique de correction ? Dans la plupart des cas, ils remplacent le mot choisi par l'élève avec un mot appartenant à un registre plus soutenu qui se distingue nettement du discours habituel de l'élève. Cette tendance avait été déjà signalée dans des études linguistiques et didactiques des années 70 (Berretta, 1977) et reconfirmée trente ans après par le linguiste A. De Benedetti

(2009). Ce dernier insiste sur l'idée que l'instruction s'est vue confier le rôle de contraster les changements dans la langue et de mettre en place un mouvement de retour en arrière, vers le prestige et la pureté originelle de la langue italienne. Le traitement automatique de l'ensemble des annotations à l'aide du logiciel *LeTrameur* nous aidera donc à confirmer cette tendance visible chez les enseignants italiens et à considérer sa présence dans le corpus français également.

L'hétérogénéité de nos écrits scolaires et la variété des styles de correction des enseignants nous empêchent d'investiguer de manière plus détaillée ici notre corpus, notre analyse n'étant d'ailleurs qu'au début. Nous concluons en mettant l'accent sur quelques résultats qui seront successivement accompagnés de pourcentages exacts : tout d'abord, nous constatons une tendance générale de l'enseignant à apporter des corrections locales, qui portent souvent sur un mot précis et sur un aspect linguistique et normatif, indépendamment de la « posture de correction » que l'on lui attribuera. De plus, à travers la phase de transcription et d'annotation, nous remarquons que les élèves italiens ont plus de facilité à modifier la structure et le contenu dans la version finale, malgré la présence de corrections de l'enseignant dans le premier jet. Cela pourrait être lié au fait que les élèves italiens soient moins inhibés par l'idée de commettre de nouvelles erreurs orthographiques. Le système linguistique joue un rôle important dans l'approche de l'erreur : l'erreur d'orthographe est considérée dans les deux contextes linguistiques comme étant plus grave par rapport aux répétitions dans la langue. Cependant, les difficultés orthographiques imposées par la langue écrite en français provoqueraient chez les élèves français un attachement au premier texte qui les rassure au moment du passage à la version finale et qui présente déjà un grand nombre de corrections, de suggestions et de commentaires de l'enseignant. Ce que l'on constate est une tendance à la mise au propre ou à l'abandon du premier jet, en faveur de l'écriture d'un nouveau texte entièrement différent, (peut-être dû aussi au fait que notre protocole de recueil ne prévoyait pas une deuxième correction du maître). En revanche, nous faisons l'hypothèse que les élèves italiens remanieraient davantage leur premier texte et qu'ils rajouteraient plus facilement du contenu : cela ne doit pas nous faire croire pourtant qu'ils aboutissent toujours à un texte de meilleure qualité dans la version définitive. Nous n'avons pas suivi les enseignants dans leur travail de réécriture avec les élèves pendant l'année scolaire, mais pouvoir investiguer la représentation de l'écriture des enseignants et ce qu'est une rédaction aboutie pour ces enseignants en particulier nous aiderait à éclaircir ce dernier point.

### **Références bibliographiques**

- BERRETTA, Monica (1977). *Linguistica ed educazione linguistica. Guida all'insegnamento dell'italiano*, Torino : Piccola Biblioteca Einaudi.
- BORE, Catherine (2007). Construire et exploiter des corpus de genres scolaires, Namur (Belgique) : Presses Universitaires de Namur, collection Diptyque.
- DAVID, Jacques & DOQUET, Claire (2016). Les écrits d'élèves : un corpus de référence pour le français contemporain. Congrès Mondial de Linguistique Française, Tours, France. hal-01659765f.
- DAVID, Jacques & DOQUET, Claire (2018). Collecter, interpréter, enseigner l'écriture », Repères [En ligne], 57 | 2018, mis en ligne le 30 juin 2018, consulté le 18 mai 2019. URL : <http://journals.openedition.org/reperes/1444>.
- DAVID, Jacques, DOQUET, Claire & RINCK, Fanny (2018). « *Que faire avec les écrits des élèves ?* », *Le français aujourd'hui*, n°203. <http://www.revues.armand-colin.com/lettres-langue/francais-aujourd'hui>.
- DE BENEDETTI, Andrea (2009). *Val piu la pratica. Piccola grammatica immorale della lingua italiana*, Bari : Editori Laterza.
- DOQUET, Claire (2011). L'Écriture débutante. Pratiques scripturales à l'école élémentaire. Rennes, PUR.

DOQUET, Claire, ENOIU, Vanda, FLEURY, Serge & MAZZIOTTI, Sara. (2017) Problèmes posés par la transcription et l'annotation d'écrits d'élèves. *Corpus, Bases, Corpus, Langage* - UMR 7320, Spécificités et contraintes des grands corpus de textes scolaire : problèmes de transcription, d'annotation et de traitement. *Corpus*. hal-01659752.

PILOGE, Jean-Luc (2010). Un lieu de tension entre posture de lecteur et posture de correcteur : les traces des enseignants de français sur les copies des élèves, *Pratiques*, 145-146 : pages 85-103.

WOLFARTH, Claire (2015). Apport du TAL à la constitution et l'exploitation d'un corpus scolaire de cours préparatoire. *Linguistique*. 2015. dumas-01167286.

### ***Références aux logiciels***

FLEURY, Serge, Le Trameur, Téléchargement et description, <http://www.tal.univ-paris3.fr/trameur/>

FLEURY, Serge. iTrameur <http://www.tal.univ-paris3.fr/trameur/iTrameur/>

FLEURY, Serge, mKAlign, Téléchargement et description, <http://www.tal.univ-paris3.fr/mkAlign/>

### ***Références aux sites Internet***

Projet E-CALM, ANR, <http://e-calm.huma-num.fr/le-projet/>

Projet ECRISCOL, Sorbonne Nouvelle-Paris 3, <http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/>

# ÉTUDE DES FORMULES EXPRESSIVES DES INTERACTIONS DANS UN CORPUS DE TWEETS<sup>1</sup>

Najwa GHARBI<sup>a</sup>

[najwa.gharbi@univ-grenoble-alpes.fr](mailto:najwa.gharbi@univ-grenoble-alpes.fr)

<sup>a</sup>LLTA Faculté des Lettres et Sciences Humaines, Sfax, Tunisie

<sup>a</sup>LIDIELM, Université Grenoble Alpes, Grenoble cedex 9, France

## 1. Introduction

Les formules expressives des interactions constituent une sous-classe des phraséologismes pragmatiques. Ce sont des expressions très fréquentes dans les interactions, comme *ça va pas la tête, c'est le comble, ça alors, il y a pas de quoi*, et qu'on désigne par d'innombrables appellations comme : « énoncés liés » (Fónagy 1982, Marque-Pucheu 2007), « routines conversationnelles » (Klein & Lamiroy 2011), « énoncés usuels » (Martins-Baltar 1997), « structures figées de la conversation » (Bidaud 2002), « phraséologismes communicationnels » (Burger 2010), « phraséologismes pragmatiques » (Dziadkiewicz 2009), « actes de langage stéréotypés » (Kauffer 2011), « phrases préfabriquées des interactions » (Agnès Tutin 2019) et bien d'autres encore. Certaines d'entre elles sont identifiées sous le terme de « pragmatème » (Mel'čuk 2013), (Fléchon et al., 2012) et (Blanco et Mejri 2018), par exemple *chapeau bas* qui est une forme type pour exprimer l'admiration, *il y pas de quoi*, une forme conventionnelle vient en réponse à un remerciement. Ces formules sont classées selon la typologie des phrasèmes proposée par Mel'čuk (2013) comme des phrasèmes pragmatiques étroitement liés à des situations d'énonciation spécifiques.

Dans le présent article, nous proposons d'étudier un éventail des formules à valeur expressive et dont la fonction est essentiellement pragmatique. Nous présenterons les propriétés formelles, syntaxiques, pragmatiques et sémantiques d'un sous-ensemble de formules expressives des interactions (désormais FEI). Ensuite, nous proposerons une grille d'analyse et nous analyserons quelques cas de FEI. Nous nous intéresserons aux actes de langage réalisés à travers les FEI. Notre analyse se base sur un corpus du réseau *Twitter*.

## 2. Choix du terme

Comme nous l'avons vu, il n'y a pas de consensus terminologique pour délimiter cette notion. Cependant, les travaux mentionnés traitent à peu près le même phénomène sous différentes dénominations. Étant donné ce flottement terminologique, nous avons choisi d'utiliser le terme « formules expressives des interactions »<sup>2</sup> pour dénommer un sous-ensemble des phraséologismes pragmatiques à fonction expressive au sens des énoncés liés de Fónagy (1928) ainsi que ceux de Pucheu (2007) et des formules de la conversation de Bidaud (2002)

<sup>1</sup> Nous remercions tout particulièrement Madame Agnès Tutin pour ses relectures attentives, ses conseils et sa disponibilité. Nous exprimons également tous nos remerciements à Monsieur Mohamed Bouattour.

<sup>2</sup> Nous avons utilisé le terme pragmatème de Gharbi (2018). Cependant, nous avons pris conscience que les critères définitoires des pragmatèmes ne correspondent pas parfaitement aux expressions étudiées qui dérivent dans bien des cas des critères principaux établis par Mel'čuk (2011). Vu le différend qu'illustre cette notion, nous avons opté pour un autre terme en collaboration avec les collègues du projet Polonium (Pragmalex, 2018/2019).

dans son ouvrage *Structures figées de la conversation: analyse contrastive français-italien*.<sup>3</sup> Le terme choisi désigne un large groupe de formules, très fréquentes dans les interactions quotidiennes écrites ou orales, notamment celles qui interviennent sur les réseaux sociaux tel que le réseau *Twitter*. Ces expressions ne s'interprètent que dans un cadre pragmatique, en faisant référence à une situation d'énonciation. Les FEI constituent le plus souvent des énoncés entiers, mais elles peuvent être intégrées dans un autre énoncé par juxtaposition ou par subordination. Les expressions choisies pour l'étude sont sémantiquement non-compositionnelles. Ce critère est beaucoup discuté, car les études sur les pragmatèmes par exemple ont classé ces expressions comme des phrasèmes compositionnelles (Blanco, 2015). C'est même un critère définitoire pour (Fléchon et al., 2012).

Autrement dit, l'utilisation de ces énoncés nécessite d'en comprendre l'emploi en contexte. Prenons par exemple l'expression *ça craint*, une expression familière qui n'a pas un rapport direct avec le sens courant du verbe *craindre* ce que pourrait être une acception du verbe *craindre*. Elle est plutôt utilisée pour exprimer plusieurs significations négatives selon les contextes situationnels dans lesquels elle a été employée. Elle peut signaler un danger possible ou une situation épineuse, voire simplement un souci. Elle peut mentionner plus encore le découragement et l'ennui face à un événement.

Nous nous intéressons à deux dimensions fondamentales : la question des actes de langage réalisés à travers les formules étudiées, correspondant à leurs fonctions pragmatiques dans le discours, et la dimension affective et émotionnelle qui les accompagne. Les deux plans doivent être soigneusement distingués, même s'ils sont étroitement liés : par exemple, la menace est un acte de langage (Weill, 1993) même si les formules de menace revêtent souvent d'une dimension affective de la colère ou de la haine.

La définition qu'illustre le mieux la notion de *formule expressive* est celle de Charles Bally<sup>4</sup> (1909) plus particulièrement dans son ouvrage *Traité de Stylistique*. Dans ce cadre, Bally a proposé la notion de « locutions exclamatives » qui servent des « moyens indirects d'expression », accompagnées toujours d'une intonation expressive. Elles se prononcent pour exprimer un sentiment ou une émotion et leur sens se détermine par le contexte. Au niveau sémantique, les locutions exclamatives sont souvent non-compositionnelles, c'est-à-dire des expressions idiomatiques comme *la belle affaire, que voulez-vous ?*<sup>5</sup> Cette étude de Bally (1909) a inspiré de nombreuses recherches qui portent sur la phraséologie pragmatique comme

<sup>3</sup> Bidaud (2002), dans son ouvrage *Structures figées de la conversation : analyse contrastive français-italien*, a relevé les formules de la conversation les plus fréquentes dans les échanges quotidiens, puis elle les a regroupées selon des étiquettes sémantiques qu'elles permettent d'exprimer un sentiment ou une sensation comme la peur, l'inquiétude, l'encouragement, la tolérance, etc. Ces étiquettes sémantiques sont considérées pour nous comme une ressource fructueuse à partir de laquelle nous avons relevé un bon nombre des formules de conversation expressives. L'ouvrage rassemble dans l'ensemble plus de 1000 expressions de la conversation.

<sup>4</sup> Charles Bally (1909) dans son ouvrage *Traité de Stylistique* a jeté les prémices du phénomène de la phraséologie française, où il propose une typologie détaillée des locutions phraséologiques, en distinguant trois types de locutions : les groupements usuels ou bien « les séries phraséologiques » et « les unités phraséologiques » et « les locutions exclamatives ». Bien que la définition de Bally soit relativement large et étendue, elle fournit des éléments très importants. Il va de soi que l'étude de la phraséologie exclamative est cruciale dans l'étude des expressions liées aux contextes, c'est-à-dire les expressions à fonction pragmatique qui ont d'innombrables appellations. Ainsi, la phraséologie exclamative sert d'appui à notre étude des FEI, car elle aborde plusieurs éléments qui peuvent nous servir dans le travail entrepris, à savoir le sens idiomatique, l'importance de l'environnement contextuel dans la production du sens, la possibilité de faire des transformations morphologiques.

<sup>5</sup> Ces exemples sont cités par Bally (seconde édition 1921 : 276).



celle de Fónagy (1982) qui a repris la notion en utilisant le terme « énoncés liés ». Ces énoncés liés répondent au stimulus déclenchant leur utilisation. C'est le stimulus qui provoque le locuteur à employer un certain énoncé en réaction ou en réponse à un autre énoncé ou face à un certain fait ou un événement. Par conséquent, un énoncé est particulièrement lié à une seule situation typique et il est contraint pragmatiquement. Nous rejoignons Fónagy (1982) pour considérer que les FEI sont étroitement liées à un contexte situationnel. Néanmoins, elles ne viennent pas toujours comme une réponse à un stimulus des interactions. De même, elles peuvent avoir plusieurs sens, tout dépend bien sûr du type d'acte auquel on a affaire, et de la situation de communication.

Nous avons essayé de choisir un terme approprié et capable de caractériser cette notion, Ce terme a été choisi en collaboration avec les collègues du projet Polonium (Pragmalex)<sup>6</sup>.

### 3. Délimitation de l'objet d'étude

La complexité de la problématique envisagée nécessite sans doute de restreindre le champ d'exploration. Devant la vaste étendue du terrain des FEI, nous avons pensé à restreindre le champ sémantique sur lequel nous effectuons cette étude, en sélectionnant le champ des affects. De ce fait, nous n'allons envisager que les expressions qui expriment des sentiments, des humeurs, des états psychologiques ou des attitudes affectives, ou une évaluation positive ou négative. Cela amène à exclure les formules de politesse et de réclamations, les salutations et tout type d'énoncés ritualisés. Nous éliminons également les interjections malgré leur valeur pragmatique et émotive, car celles-ci nécessitent une étude à part entière du fait de leur étendue. Il est à noter que les interjections ont été traitées notamment par Bally (1909) et Blanco & Mejri (2018).

En outre, notre étude se consacrera seulement aux expressions polylexicales, car le critère de la polylexicalité constitue un caractère fréquent pour les FEI. Nous allons donc éliminer les expressions qui se composent d'une seule unité lexicale, et nous nous intéressons donc aux formules qui constituent des unités minimales communicatives (Zifonun et al., 1997 : 87). Les formules monolexicales sont souvent étudiées comme marqueurs discursifs (Dostie, 2004).

D'un point de vue sémantique, les expressions que nous aborderons sont sémantiquement non-compositionnelles, car elles sont composées de plusieurs mots et possèdent un sens global qu'on ne peut pas décomposer à partir de chaque mot. Cela présuppose qu'elles sont gravées dans la mémoire discursive des locuteurs et des interlocuteurs concernés.

### 4. Constitution du corpus

Ce travail est mené à partir d'un corpus de tweets informatisé du réseau *Twitter*. Il s'agit d'un corpus déjà constitué dans le cadre du projet CoMeRe<sup>7</sup> (communication médiée par les réseaux). CoMeRe s'inscrit dans le cadre du projet « Humanités numériques et data-journalisme : le cas du lexique politique ». Le corpus a été élaboré par plusieurs chercheurs de l'université de Cergy Pontoise et de l'équipe MIDI, à savoir Julien Longhi, Claudia Maricina, Boris Borzic, Abdulhafiz Alkhoul. Il vise à rassembler différents corpus représentatifs des

<sup>6</sup> Le projet POLONIUM 2018-2019 (« Pragmatèmes en contraste : de la modélisation linguistique au codage lexicographique ») est un projet franco-polonais financé en Pologne par le Ministère de la Science et de l'Enseignement supérieur (MNiSW) et en France par les ministères des Affaires étrangères et du Développement international (MAEDI) et de l'Éducation nationale de l'Enseignement supérieur et de la Recherche (MENESR). Il est coordonné par Francis Grossmann (Université Grenoble Alpes, Lidilem, France, Grenoble) et Anna Krzyżanowska (Université Marie Curie-Skłodowska, Institut de Phonologie Romane (Pologne, Lublin).

<sup>7</sup> Tous les corpus sont disponibles et téléchargeables sur le site du projet <http://comere.org> ou sur le site d'Ortolang.

formes de communication sur les réseaux tels que les tweets, les blogs, les forums, les courriers électroniques. Ce projet a contribué à créer un noyau de corpus de communication médiée par les réseaux (Computer Mediated Communication – CMC). Tous les corpus CoMeRe sont accessibles et téléchargeables en ligne sur le site d’Ortolang<sup>8</sup>. Le projet met à notre disposition des corpus européens de nouveaux moyens de communication en trois langues : a) corpus allemand de type CoMeRe/CMC : projet DeriK, Corpus allemand DWDS b) corpus hollandais/flamand SoNaR, c) corpus français<sup>9</sup>.

Nous nous intéressons particulièrement au corpus français qui comprend plusieurs genres de communications médiées par le réseau comme les chats, les discussions sur des forums, les SMS, et les emails. Le tableau suivant détaille les différents types de corpus français au sein du projet CoMeRe<sup>10</sup>.

<p><b>SMS</b></p> <ul style="list-style-type: none"> <li>- <a href="#"><u>cmr-smslareunion</u></a></li> <li>- <a href="#"><u>cmr-smsalpes</u></a></li> <li>- <a href="#"><u>cmr-88milsms</u></a></li> </ul> <p><b>Wiki discussions</b></p> <ul style="list-style-type: none"> <li>- <a href="#"><u>cmr-wikiconflits</u></a></li> </ul>	<p><b>Tweets</b></p> <ul style="list-style-type: none"> <li>- <a href="#"><u>cmr-polititweets</u></a></li> <li>- <a href="#"><u>cmr-intermittent</u></a></li> </ul> <p><b>Weblog</b></p> <ul style="list-style-type: none"> <li>- <a href="#"><u>cmr-infral</u></a></li> </ul>	<p><b>Email</b></p> <ul style="list-style-type: none"> <li>- <a href="#"><u>cmr-simuligne</u></a></li> </ul> <p><b>Discussion forum</b></p> <ul style="list-style-type: none"> <li>- <a href="#"><u>cmr-simuligne</u></a></li> </ul>	<p><b>Text chat</b></p> <ul style="list-style-type: none"> <li>- <a href="#"><u>cmr-getalp_org</u></a></li> <li>- <a href="#"><u>cmr-favi</u></a></li> <li>- <a href="#"><u>cmr-favi (POS tagged)</u></a></li> <li>- <a href="#"><u>cmr-simuligne</u></a></li> </ul>	<p><b>Multimodal</b></p> <ul style="list-style-type: none"> <li>- <a href="#"><u>cmr-copeas</u></a></li> <li>- <a href="#"><u>cmr-tridem06</u></a></li> </ul> <p><b>Multimodal + 3D</b></p> <ul style="list-style-type: none"> <li>- <a href="#"><u>cmr-archi21</u></a></li> </ul>
--	--	--	--	--

**Tableau 1 - Les genres de communications médiées dans le projet CoMeRe (CMC)**

Parmi ces corpus, nous avons opté pour le sous-corpus « Intermittent »<sup>11</sup> qui constitue la totalité des tweets de 215 utilisateurs actifs sur le thème des intermittents en France lors d’une controverse à leur sujet de juin à septembre 2014. Les comptes de personnalités ont permis l’extraction de 586 276 tweets. Nous avons également exploité le corpus Polititweets<sup>12</sup> qui provient de comptes politiques influents dans la campagne des municipales de mars 2014. Notre corpus compte donc deux sous-corpus, répartis comme suit :

<sup>8</sup> Les corpus sont mis en format TEI (Text Encoding initiative) où tous les éléments sont codés dans le but de baliser tous les aspects technologiques propres au réseau *Twitter* à savoir le hashtag, la mention par l’arobase (@). (Longhi, 2013)

<sup>9</sup> <https://repository.ortolang.fr/api/content/comere/v3.3/comere.html>

<sup>10</sup> Le tableau ci-dessus est tiré du site du projet CoMeRe, cité précédemment.

<sup>11</sup> Les intermittents du spectacle en France sont les artistes et les techniciens qui travaillent en intermittence pour des entreprises du spectacle et dans le domaine du cinéma, et de l’audiovisuel. Ils bénéficient d’une assurance chômage, car ils ne sont pas des employés permanents. En 2014, le Medef a demandé de supprimer ce régime, ce qui a enflammé les discussions autour ce thème controversé de juin à septembre 2014.

<sup>12</sup> « Polititweets » s’insère dans un contexte révélateur, celui de la campagne des municipales de mars 2014. Or, les tweets proviennent des comptes politiques les plus influents, extraits des comptes de 7 personnalités politiques, JL Melenchon, F Bayrou, Jean-François Copé, Marine Le Pen, Jean-Marc Ayrault et Daniel Cohn- Bendit. Il permet de sélectionner la période d’entre-deux élections municipales 2014. Les comptes sont en nombre de 205, ils ont permis d’extraire 34273 tweets au total.

Corpus	Nombre de twittos	Nombre de tweets	Nombre de mots
Polititweets	205	586 276	91 765
Intermittents	215	34 273	14 203
Total	420	620 549	105 968

Tableau 2 - Récapitulatif du corpus de tweets en chiffre<sup>13</sup>

Le travail sur un corpus de tweets peut paraître audacieux, car ce type de corpus pose généralement maintes difficultés de délimitation. Tout d'abord, ce type de corpus reste difficile à cerner en raison de la quantité des données diffusées. De toute évidence, n'importe qui peut poster un message, un commentaire sur *Twitter*, à la simple condition d'y ouvrir un compte d'utilisateur. Il faut noter également que la grande majorité des tweets sont diffusés publiquement, et ils peuvent donc être consultables par tous les twittos. De ce fait, afin d'obtenir une certaine cohérence, il faut délimiter les corpus soit pour une période déterminée, soit par un évènement précis. Les concepteurs du projet ont pris conscience de cette difficulté et ils ont choisi une manière efficace pour extraire les tweets.

La construction du corpus intermittent a été effectuée à partir du hashtag « #intermittent ». Nous avons au total 13 074 tweets avec ce hashtag dans la période indiquée dans le corpus. Ce corpus a été construit à partir de 4617 twittos. Dans ce cadre, Paveau (2013 : 9) a considéré le hashtag comme « un technomorphème » qui fonctionne comme un organisateur des discours en ligne, et qui permet la détection d'évènements et nous donne accès à un fil.

Le corpus « Polititweets » est constitué d'une façon moins compliquée que le premier. Les tweets qui constituent l'ensemble du corpus sont extraits des comptes politiques des candidats des élections municipales de 2014. Il ne s'agit pas de collecter un ensemble représentatif d'échanges des personnalités politiques, c'est-à-dire des échantillons à observer, mais plutôt de la totalité de tweets postés dans la période examinée. Au vu de la difficulté de construire ce type de corpus, nous avons utilisé un corpus déjà construit dans le cadre d'un projet ce qui a grandement facilité notre travail.

## 5. Spécificités du corpus étudié

Avant de justifier notre choix de corpus, nous allons donner une description du réseau Twitter duquel provient notre corpus. *Twitter* est un réseau social numérique créée en 2006. Il permet d'écrire, de partager et de commenter des contenus de communication. Cette plateforme permet de communiquer des opinions portant souvent sur des évènements ou des personnes, et d'exprimer des sentiments, de chercher à susciter les interactions d'autres « twittos » (utilisateurs de Twitter). Par ailleurs, la taille d'un tweet est restreinte à 140 caractères. Cette contrainte qui pèse sur le nombre de caractères employés pour « tweeter » (écrire un tweet) exige, d'un côté, une expression brève et concise, mais d'un autre côté, elle favorise plus souvent l'utilisation des expressions et des formules productives comme le montre Longhi : « Sur le plan de la forme des tweets, la taille réduite des productions se prête particulièrement bien à la production de « petites phrases » ou formules » (2014 : 2).

Pankhurst (2009) a évoqué les points communs entre les tweets et les messages SMS, dans la mesure où les deux formes tendent à la brièveté du discours électronique médiée (DEM). De surcroît, cette forme courte fait appel à la concision d'écriture sur le plan discursif. L'utilisation des formules expressives peut être une solution pour gérer un espace d'écriture restreint, comme le souligne Longhi : « (...) il véhicule des formes qui peuvent devenir de petites

<sup>13</sup> Il y a plus de tweets que de mots dans notre corpus, cela est dû à une opération de dédoublement (suppression des chaînes retweetées).

phrases sur le plan discursif, et s'accompagne d'une intensité sémantique originale » (2014 : 11).

Par conséquent, la fameuse contrainte formelle des tweets (140 caractères maximum)<sup>14</sup> implique le plus souvent la possibilité d'utiliser des formules ou des messages brefs.

Le choix de ce corpus n'est pas le fait du hasard. Nous avons choisi les tweets comme objet de notre étude pour plusieurs raisons. Notre choix est motivé par l'intérêt porté aux communications médiées par les réseaux dans les recherches en sciences du langage. Dans les dernières années, l'usage des réseaux sociaux joue de plus en plus un rôle central dans la vie quotidienne des gens et peu à peu, ils sont devenus des formes de communication les plus fréquentes. Il y a de toute évidence des spécificités particulières à ce type de corpus par rapport aux autres types de corpus littéraires ou journalistiques. De ce fait, les tweets constituent un corpus très intéressant encore assez peu étudié jusqu'à présent.

Commençons par le choix du sous-corpus « Intermittent » qui se situe dans une période de controverse, celui de la polémique sur le statut des intermittents en France. Le choix de la thématique nous a permis d'effectuer un travail de recherche sur des productions discursives réelles et spontanées, spontanéité qui n'est pas forcément présente dans les tweets politiques. Les tweets ont été considérés par les intermittents du spectacle non pas comme un canal de transmission d'informations seulement, mais aussi comme un moyen de s'exprimer. Tous les tweets contenant le hashtag intermittent de juin à septembre 2014 s'inscrivent dans le refus de la suppression de ce régime de travail qui s'exprime par des expressions qui soulignent des sentiments souvent négatifs comme la colère, l'indignation, la menace et le regret.

Dans le sous-corpus « Polititweets », qui comprend des interactions authentiques de discours politiques, la majorité des tweets portent sur des interactions autour d'événements de l'actualité ou sur les jugements des personnes. Ils ont servi à exprimer des opinions politiques qui s'accompagnent souvent des jugements évaluatifs positifs ou négatifs. Nous avons opté pour ce corpus pour observer comment les FEI fonctionnent dans les interactions politiques. Par ailleurs, nous sommes consciente que ces tweets sont loin d'être spontanés, chaque commentaire partagé, tweet posté ou position diffusée sont bien calculés et étudiés.

Il va sans dire que les personnalités politiques n'exploitent pas cet espace scriptural de manière irréfléchie. Elles sont bien au contraire très vigilantes quant à l'effet produit. Nos corpus sont complémentaires : un spontané, l'autre beaucoup plus surveillé. Nous avons décidé d'évoquer le discours politique dans des environnements numériques pour examiner la grande variété et la richesse discursive des interactions politiques, et afin de dégager les spécificités des lexiques politiques sur le plan syntaxique, morphologique et pragmatique.

Pour résumer, les tweets se présentent comme des communications instantanées d'une idée, d'un sentiment, d'un jugement des communications de l'instant, une idée, un sentiment à exprimer, un jugement négatif ou positif à communiquer (Gharbi 2018 : 155).

### ***Les données du corpus***

Dans la présente section, nous allons brosser les étapes qui ont précédé l'extraction de données du corpus. La tâche a consisté à suivre les étapes suivantes :

- L'élaboration d'une liste des formules expressives des interactions (prétraitement)

<sup>14</sup> En novembre 2017, le réseau Twitter a doublé le nombre de caractères pour les messages sur le micro-blogging pour atteindre 280 caractères. Comme notre corpus est constitué entre 2013 et 2014, il n'est concerné par ce processus de doublement.

Pour rassembler une liste des FEI, nous avons exploité tout d'abord des ressources dictionnairiques telles que *Le Grand Robert*<sup>15</sup>, le *TLFI*<sup>16</sup>, *wikitionnaire*<sup>17</sup> et le Dictionnaire spécialisé *d'Expressions et Locutions* de Sophie Chantreau et Alain Rey (2005)<sup>18</sup>. Ce dernier a constitué pour notre travail une ressource très intéressante, car il contient un grand nombre de FEI, bien qu'elles ne soient pas décrites à partir de leurs entrées propres. Nous nous sommes appuyée aussi sur les travaux relatifs à la notion de phraséologismes pragmatiques. Parmi les plus marquants, citons le travail de Bidaud (2002) dans son ouvrage *Structures figées de la conversation : analyse contrastive français-italien*. Bidaud a proposé un recensement de 1400 formules conversationnelles du français et leurs équivalents en italien. Cet ouvrage a constitué pour nous une ressource précieuse, car toutes les expressions sont associées à des étiquettes sémantiques bien détaillées, commentées et illustrées dans des citations. Bidaud (2002) a fourni un recueil d'expressions français-italien qui sert à exprimer les attitudes et l'expression des sentiments. On y trouve des façons de refuser, d'exprimer l'accord ou le désaccord, la certitude ou le doute, porter un jugement ou exprimer un sentiment. On s'intéressera bien entendu à celles qui véhiculent un sens expressif comme la surprise, la colère, la peur et nous avons exclu les structures qui expriment par exemple, la certitude, l'affirmation, l'accord, etc.

Ensuite, nous avons recouru aux travaux sur la notion de FEI, notamment les travaux de Kauffer (2011 à 2017), de Blanco et Mejri dans leur ouvrage collectif *Les Pragmatèmes* (2018), et de Klein et Lamiroy (2006, 2011). Nous nous sommes également servi de deux fichiers *Excel* élaborés dans le laboratoire <sup>19</sup>*de Linguistique et Didactique Des Langues Etrangères et Maternelles (LIDILEM)* par Agnès Tutin à partir d'une liste établie par D. Le Pesant (à paraître). Le premier fichier comprend 2081 expressions phrastiques et normalisées à partir desquelles nous avons relevé les expressions à sens expressif. De plus, nous avons bénéficié d'un second fichier qui englobe 1048 énoncés expressifs, annotés et arrangés sur le tableau (*Excel*). Ainsi, nous avons créé une liste qui englobe près de 1000 formules expressives.

Nous avons par la suite codé cette liste sur le tableau (*Excel*), où nous avons indiqué la définition de chaque expression, comme le montre le tableau suivant.

---

<sup>15</sup> *Grand Robert*, Version électronique.

<sup>16</sup> <http://atilf.atilf.fr/>

<sup>17</sup> <https://fr.wiktionary.org/>

<sup>18</sup> *Dictionnaire des Expressions et Locutions*, Alain Rey & Sophie Chantreau (2005).

<sup>19</sup> Laboratoire de Linguistique et Didactique Des Langues Etrangères et Maternelles (LIDILEM) Grenoble- France.

	A	B	C	D	E	F
1	Termes	Type	Registre	Statut syntaxique	Définition	Source
2	à la bonne heure	prag	familier	PA	C'est très bien, c'est parfait, tant mieux exprime l'assentiment, l'approbation, parfois d'une manière ironique par antiphrase, s'emploie familièrement au sens de au 'moment propioe'	Rey, Chantreau
103	c'est le complet	prag	familier	PV	il ne manquait que cela	Rey, Chantreau
104	c'est le pompon	prag	familier	PV	c'est le comble	Rey, Chantreau
105	c'est ma mort	prag	familier	PV	c'est terrible	Rey, Chantreau
106	c'est marre	prag	familier	PV	ça suffit	Rey, Chantreau
107	c'est n'importe quoi	prag	courant	PV	Un objet ou une idée dont la précision importe peu et que l'on peut choisir librement	wiktionnaire
108	c'est pas cochon	prag		PV	ce n'est pas mal	Rey, Chantreau
109	c'est pas de la tarte	prag	familier	PV	ce n'est pas facile	dictionnaire reverso
110	c'est pas dommage	prag	courant	PV	enfin! S'emploie pour signifier que la chose qui vient d'arriver était attendue depuis longtemps	Rey, Chantreau

Tableau 3 - Liste des FEI

Dans le tableau ci-dessus, nous avons extrait les formules des interactions qui expriment des affects. Nous avons cherché la définition de chaque expression dans une ressource lexicographique, ainsi que son type lexical. La création d'une liste constitue une étape fondamentale du travail et elle constitue un préalable à tout traitement. Néanmoins, la difficulté majeure qui se pose pour comme dans toute recherche phraséologique est d'arriver à une liste exhaustive et définitive. Le recueil d'une liste exhaustive constitue un horizon inatteignable.

- Recherche des concordances

Nous avons cherché des exemples dans des concordances via l'outil de traitement de corpus AntConc<sup>20</sup>. Il s'agit d'un logiciel gratuit et téléchargeable en ligne à partir du site mentionné ci-dessus. Il est créé par Laurence Antony et il permet d'effectuer une recherche d'un mot, d'expressions ou des listes de mots dans un ou plusieurs textes. Il est à noter qu'une relecture manuelle intégrale du corpus a été effectuée pour obtenir des résultats exhaustifs. Ce processus nous a permis de trouver des expressions qui figuraient dans notre liste. La relecture s'est avérée nécessaire. C'est pour cela que nous pouvons affirmer que le travail de traitement n'est pas complètement automatique et que nous sommes intervenue manuellement à plusieurs reprises.

À l'issue de cette étape, nous traitons nos données dans un fichier *Excel*, dans lequel nous avons précisé le type sémantique de chaque expression relevée comme le montre l'illustration suivante, tirée du corpus de tweets :

<sup>20</sup> <http://www.antlab.sci.waseda.ac.jp/>

1	2	3	4	5	6	7	8	
Expression	Registre	Type sémantique	Définition	Source	Type	Politiwe ets	Intermitt ent	
plus jamais ça!	Familier	refus, colère	refus du pessimisme	liste Bidaud	polylexical		1	0 plus jamais ça , s
pourquoi donc?	Familier	sens ironique	l'adverbe "pourquoi" exprime pour <u>quelle</u> raison ; dans	Wiktionnaire	polylexical			
putain de bordel de merde!	vulgaire	ressentiment	juron	Grand Robert	polylexical		1	0 Putain de bordel de mal finir... #inter
quel bordel?	vulgaire	l'emportement	juron	Grand Robert	polylexical		0	1 Quel bordel! Ce n'
ras le bol!	très familier	irritation, exaspération	Mécontentement, exaspération	Wiktionnaire	polylexical		1	0 <p>Ras le bol des d du FN,
rejoignez-vous	courant	encouragement	inviatation à rejoindre, incitation affective	Notre corpus	polylexical			3
rien de pire	Familier	satisfaction	situation mauvaise, synonyme de c'est un	Notre corpus	polylexical		0	1 <p> .@frehsamen Ac son duché, rien de
rien moins que ça!	Familier	stupéfaction	c'est un comble	liste Bidaud	polylexical		1	0 leur auto-promotic scientifiques
rigole plus!	Familier	avertissement	exprime la menace	TLF	polylexical		0	1 rigole plus

Tableau 4 - Extrait des concordances à partir d'AntConc

## 6. Résultats

Au total, 79 formules expressives ont été repérées dans le corpus, dont 43 dans le sous-corpus « Polititweets » et 36 dans le sous-corpus « Intermittent », ce qui représente une proportion comparable entre les deux corpus. Les formules relevées sont assez variées. Elles sont au niveau grammatical pour la majorité des phrases phrastiques (phrases tensées) *tiens donc*, *c'est bon*, *ça va !* (67%) d'occurrences dans « Polititweets » et (58%) dans « Intermittent ». Les autres occurrences rencontrées constituent des locutions interjectives ou des jurons *Bon Dieu*, *Nom de Dieu*.

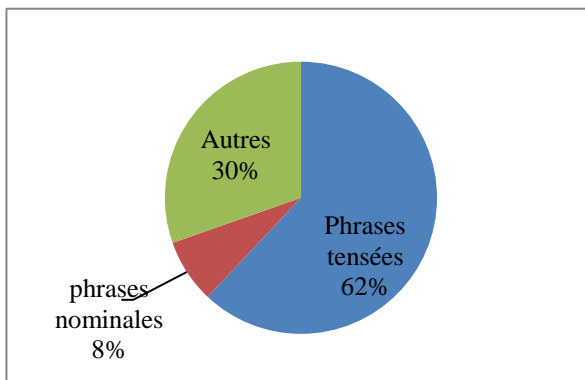


Figure 1 - Répartition des FEI selon le type grammatical

Il est évident que le corpus apparaît assez modeste pour mettre en évidence tous les phénomènes représentatifs de ce fait linguistique. Toutefois, on ne peut pas masquer la variété des expressions rencontrées. Le grand avantage de l'étude des tweets est leur nature authentique et spontanée. Ce type de corpus présente des interactions familières dans des contextes quotidiens. Au niveau discursif, les FEI fonctionnent comme un moyen pour véhiculer des émotions et effectuer des actes de langage réactifs et expressifs. Nous nous sommes appuyée sur des exemples attestés et authentiques, mais ils restent peu suffisants par rapport aux travaux menés sur les corpus littéraires.

## 7. Propriétés des formules expressives des interactions (FEI)

A partir des expressions repérées dans notre corpus, nous avons observé un ensemble de propriétés.

### 7.1. Des formules partiellement fixes

#### 7.1.1. Insertion

Il est communément admis que les phraséologismes pragmatiques ne sont pas des syntagmes libres sur le plan lexical, c'est-à-dire que le locuteur n'a pas le choix des unités lexicales qui les composent. C'est pour cette raison qu'on les retrouve souvent classées comme des types d'expressions figées. Nous entendons par fixation lexicale l'impossibilité d'établir des alternances paradigmatiques et les insertions au plan syntagmatique. Néanmoins, ce critère n'est pas complètement établi, car il y a des expressions qui acceptent l'insertion syntagmatique d'un élément facultatif par exemple *c'est fort de café* qu'on utilise pour manifester la colère ou l'indignation peut être employé comme suivant *c'est un peu fort du café*, ou *c'est parti* vs. *c'est reparti*. Fónagy (1997), Marque-Pucheu (2007) et Klein & Lamiroy (2011) observent aussi des variations comme *ferme ta gueule !* qui peut varier sur l'axe paradigmatique *ferme ton clapet*<sup>21</sup>. Sur le plan lexical, on peut observer quelques variations distributionnelles et paradigmatiques.

#### 7.1.2. Caractère elliptique

Ce caractère consiste à supprimer d'un énoncé un ou plusieurs éléments lexicaux. Les twittos recourent à un langage elliptique, bref et très réactif. Ce processus de réduction donne lieu à des formules variées. L'ellipse se révèle dans la chute de certains éléments lexicaux par exemple l'effacement soit du déterminant démonstratif, soit de la suite du verbe *être* et l'adverbe de négation comme *ce n'est pas possible* vs. *c'est pas possible*, voire *pas possible* qui s'utilise pour mettre en scène plusieurs significations selon le contexte, comme l'étonnement, la colère. *Ce n'est pas du jeu* vs. *c'est pas du jeu*, est souvent utilisé pour protester ; *ce n'est pas la joie* vs. *pas la joie* qui s'emploie par le locuteur pour dire que la situation est désagréable, pénible, voire triste. Conséquemment, les tweets intègrent fréquemment l'effacement de certains constituants de l'énoncé pour des fins liées au laconisme qui peut également être dû à des interactions émotionnelles et immédiates. Le fait que les FEI montrent des variations lexicales n'empêche pas la présence des expressions dont le figement est intégral comme *nom de Dieu, tant pis, ça suffit*.

### 7.2. Préfabrication et non-compositionnalité

Les formules expressives des interactions sont souvent sémantiquement non-compositionnelles, c'est-à-dire que leur sens global ne se construit pas à partir leurs constituants lexicaux. Dans ce cas, les constituants lexicaux ne suffisent pas pour comprendre

<sup>21</sup> Bouche/caquet/boîte à camembert ; il y en a beaucoup !



ce que la formule veut dire. Prenons l'exemple de l'expression *il n'en est pas question* utilisée souvent par ellipse sous la forme *pas question* qui exprime un refus affectif lié à un état émotionnel d'exaspération ou d'extrême colère et peut être synonyme des formules comme *des clous, des nèfles*. Ici, nous ne recourons pas au sens du mot *question* pour comprendre le sens global de la formule *pas question*. Ceci est une conséquence de la non-compositionnalité sémantique.

Nous ajoutons l'exemple de *ça suffit* qui est très courant dans les interactions. Cette formule ne signifie pas que le locuteur est content et satisfait de quelque chose. En revanche, le locuteur fait usage de cette formule pour manifester son irritation et son impatience devant une situation agaçante. Elle signifie *un ras-le-bol, être exaspéré, en avoir marre* de quelque chose qui va franchir les bornes que le locuteur estime raisonnables. Dans ce cas, cette formule peut être synonyme de formules comme *ta gueule, la ferme, ferme-la*, mais la formule, *ça suffit* apparaît moins violente. Examinons maintenant la formule *ça ne manquera pas de sel* dans l'énoncé suivant.

- (1) J'entends les annonces. Je veux qu'on passe d'un Président des annonces à un Président des actes. Aux vœux de @Francois Fillon a 11:30. **Ça ne manquera pas de sel.**

Cette formule est employée dans les interactions pour signifier ironiquement que la parole prononcée est drôle et cocasse. Il s'agit d'un commentaire critique destiné à l'interlocuteur pour lui dire qu'on est insatisfait et encore déçu. Elle se caractérise par une tournure idiomatique. Selon le contexte, elle véhicule un sentiment d'insatisfaction et de déplaisir ou plutôt de dérision face à une parole, une situation ou un comportement.

En bref, le sens des FEI n'est pas compréhensible littéralement, c'est-à-dire que c'est l'usage social et linguistique qui les construit et non des règles linguistiques prévisibles. En cela, elles nécessitent pour leur compréhension tout un bagage socioculturel. Elles sont bien identifiées, cela veut dire qu'elles sont reconnaissables dans une communauté linguistique déterminée. C'est pour cela qu'elles présentent souvent des difficultés majeures pour les non-natifs en raison de leur complexité sémantique.

### 7.3. Propriétés morphosyntaxiques

#### 7.3.1. Au niveau syntaxique

Les FEI se distinguent par des formules qui s'écartent parfois des formes les plus courantes. Cela veut dire qu'elles ne suivent pas toutes les normes syntaxiques. Prenons les exemples suivants :

- (2) Mathieu Grégoire : « S'attaquer aux droits des chômeurs me paraît insensé, **va pas la tête** ? regardez ce que vous avez fait »
- (3) à voir: débat républicain avec Renaud Muselier sur LCI en direct de marseille-bennahmias-vs-muselier-sur-lci-municipales2014\_news"**pas la joie** ».
- (4) si ça, **c'est pas du vol...** Précaires et intermittents en lutte, 1 arrêté à sortie occupation Hotel Lutetia, 27e H de garde à vue et risque comparution demain: scandale!
- (5) Dans un an (2015) le #Center Parcs de la #Vienne (86) ouvrira ses portes. 1000 personnes vont participer au chantier. **C'est bien parti !**

On lit à la place de *ça va pas la tête* dans l'énoncé (2) *va pas la tête* ce qui témoigne les écarts éventuels de la syntaxe normée avec l'absence de sujet. Dans l'énoncé (3), on observe une absence de sujet et de verbe copule. Dans le dernier exemple, c'est l'adverbe de la négation *ne*

qui est absent de l'énoncé. Par ailleurs, dans notre corpus, les FEI correspondent souvent à des phrases complètes dont les structures syntaxiques sont variées. Il y a des exceptions des formules qui constituent une partie d'énoncé (proposition principale ou proposition subordonnée ou complétive). Elles ne constituent pas souvent des phrases assertives comme le cas de *c'est pas du vol*, *pas la joie* dans les énoncés (3) et (4). Une formule peut aussi être une phrase interrogative comme dans *ça va pas la tête ?* qui s'utilise couramment pour souligner un état émotionnel de colère. On note ici l'absence des marques morphosyntaxiques des interrogations comme les pronoms interrogatifs. Le locuteur n'utilise pas cette phrase interrogative pour poser une question ou pour demander d'information. Elle est au contraire une affirmation déguisée où on pourra l'appeler une interrogation rhétorique. Dans l'énoncé (5), la formule correspond à une phrase *c'est bien parti !* Elle est expressive, car elle véhicule un sentiment d'enthousiasme et elle provoque une réaction chez le destinataire. Les phrases interrogatives et exclamatives sont considérées pour Bally comme « des moyens indirects d'expression » (1909 : 256). Nous rejoignons Bally pour dire que ces modalités se chargent souvent du contenu expressif ou affectif.

Dans les tweets, il n'est pas rare de trouver des formules qui omettent l'adverbe de négation ou le pronom démonstratif. Il faudra admettre que ces formules ne sont pas ni fautives ni anormales, mais qu'elles sont proches de la langue orale dans les écrits numériques sur les réseaux sociaux. Nous avons eu une énorme difficulté à décider si ces formules appartiennent ou non aux expressions figées, mais elles ne peuvent pas certes être rangées parmi les constructions canoniques non plus.

### 7.3.2. *Le niveau morphologique*

Le blocage grammatical d'une formule est constitué par l'impossibilité de modifier un trait tel que le nombre ou le temps de différentes composantes de l'expression. Dans ce qui suit, nous allons observer, à travers des exemples concrets tirés de notre corpus, le degré de figement morphologique des formules étudiées.

Pour le changement de nombre, les FEI sont strictement figées et par conséquent elles n'admettent aucun changement de nombre. Prenons l'exemple de *c'est une blague ? \* ce sont des blagues ? Ça va pas la tête ? \* ça va pas les têtes ?*

Observons maintenant les restrictions du temps dans les formules expressives des interactions. En nous appuyant sur notre corpus, il est impossible d'utiliser différentes formes verbales pour une seule expression comme *ça craint*, *ça suffit*, *allons-y*, *c'est parti* qui sont fixes au présent de l'indicatif. Chaque changement opéré au niveau du temps entraîne la modification du sens global de la formule. Cependant, on peut parfois observer des variations limitées au niveau du temps. Bidaud (2002) a bien souligné des structures de conversation sous différentes formes verbales comme *il ne manquait plus que ça*, *il ne manquerait plus que ça*. Il faut noter que ce type de variation ne figure pas dans notre corpus.

En ce qui concerne les formes verbales, il est possible qu'une formule admette des variations des formes verbales par exemple la formule *excuse-moi excusez-moi*<sup>22</sup> pour demander le pardon, plus encore la formule *voulez-vous* ou *que veux-tu* qui exprime l'impuissance. Pratiquement toutes les formes à la deuxième personne admettent ce type de variation.

On pourra donc supposer qu'on rencontre plus qu'une forme verbale. La plupart des occurrences sont soit au présent de l'indicatif soit au futur simple ou à l'imparfait. Ainsi, les possibilités de changement de nombre et de temps sont assez restreintes dans notre corpus. De

<sup>22</sup> T'inquiète pas (vous inquiétez pas), vas-y (allons-y allez-y), veux -tu ça (voulez-vous ça), ferme-la (fermez-la), ferme ta gueule (fermez vos gueules), ferme ta boîte (fermez vos boîtes), ne t'en fais pas (ne vous en faites pas), tu rigoles (vous rigolez), tiens bon (tenez bon).

cette façon, nous pourrions dire que les FEI se caractérisent par un blocage morphologique élevé.

#### 7.4. Liens avec la situation d'énonciation et les fonctions pragmatiques des FEI

Nous avons examiné jusqu'à présent plusieurs critères définatoires des FEI, mais nous n'avons pas examiné un facteur de première importance dans l'étude des FEI, celui du contexte. Elles sont souvent fixes pragmatiquement, cela veut dire qu'elles sont réservées à un type spécifique de situation. Par conséquent, elles sont dotées d'une force illocutoire et elles correspondent à des actes de langage expressifs ciblés.

- (6) Thierry Lepaon : "Le patronat est assisté par le gouvernement, **c'est un comble**" -ce sont les entreprises donc l'emploi "rassuré, serein et déterminé" à quelques jours des #Municipales2014 les indicateurs sont favorables pour sa victoire.

La formule *c'est un comble* est utilisée pour s'indigner face quelque chose de peu ordinaire. Rey et Chantreau (2005) estiment qu'elle est le synonyme de *c'est la meilleure*, *c'est la totale*, *c'est le pompon*, *c'est complet*. Elle a un sens pragmatique d'indignation. À la différence de cette formule qui est liée à un seul acte de langage, il existe des énoncés qui se rapportent à des signifiés variés. Dans ce cas, il faut examiner profondément les données contextuelles pour dégager le type d'acte. Le contexte joue ici un « rôle décisif » comme le montre Orecchioni (2005). Voici les exemples ci-dessous :

- (7) Le logement social pr nous n'est pas un gros mot! pr tous est note projet! **Allons-y!**"
- (8) Ecoutes Sarkozy prjf- cope **Trop c'est trop**, cette affaire est désormais d'une gravité exceptionnelle : c'est l'État de droit qui est en cause :
- (9) **ah bon?** Pquoi la politique, le débat social est scandaleux? Ces émissions sont fabriquées par des sal.
- (10) FILIPETTI<sup>23</sup> avertit sur BFMTV<sup>24</sup> que les intermittents ne doivent empêcher spectacles festival. **BEN VOYONS.**

La formule *allons-y*, dans l'énoncé (7), admet aussi la forme de la deuxième personne de singulier ou du pluriel *allez-y* ou *vas-y*. Le pronom *y* ne renseigne pas ici sur un lieu, mais il s'amalgame avec le verbe être conjugué à la première personne du pluriel pour signifier un cri d'encouragement ou d'enthousiasme envers un événement ou une action. C'est déjà l'explication que propose Bidaud (2002). Néanmoins, cette expression va à l'encontre des étiquettes sémantiques de Bidaud pour indiquer d'une manière ironique une critique adressée à l'interlocuteur.

La formule *trop c'est trop* souligne la protestation du locuteur devant une limite dépassée comme l'indique Bidaud (2002). L'utilisation de cette formule s'accompagne normalement d'un sentiment de mécontentement et d'emportement. Ceci nous renseigne aussi sur le type du rapport entre l'interactant qu'est ici un candidat des élections municipales de 2014 Jean-François Copé avec son destinataire. Nous avons mentionné les interactants dans cet énoncé, car majoritairement les interactions sur *Twitter* s'adressent à un destinataire collectif.

<sup>23</sup> Aurélie Filippetti, est une femme politique et romancière française. Elle fut ministre de la culture et de communication du 16 mai 2012 au 25 août 2014 dans les gouvernements Ayrault I et II, puis Valls I.

<sup>24</sup> BFM TV est une chaîne de télévision française d'information nationale.

La formule *ah bon* est très souvent utilisée dans le langage parlé pour marquer des sens variés comme l'étonnement et la surprise et bien évidemment l'incrédulité et elle renvoie davantage le sentiment de méfiance à l'égard des propos d'autrui. Dans cet énoncé, le locuteur veut réaliser un acte d'incrédulité.

Dans le dernier énoncé, *ben voyons* sert d'ordinaire à marquer ironiquement l'incrédulité. Elle s'utilise souvent pour mettre en doute l'affirmation de l'interlocuteur et pour dire qu'on n'est pas d'accord avec lui. Dans ce sens, elle peut être l'équivalente de formules comme *mon œil, tu parles*. Toutefois, le sens peut être dissimilé dans l'interaction et il n'est pas compréhensible immédiatement. Dans notre exemple, le locuteur vient d'exprimer son sentiment d'indignation à l'égard du propos de l'interlocuteur. On comprend aussi qu'une critique se dissimule derrière cette formule. Le sens de cette formule est aussi ambigu dans la communication et il n'apparaît qu'en contexte d'énonciation.

Pour résumer, nous dirons que les FEI sont plus souvent monofonctionnelles. Cela veut dire qu'elles possèdent une valeur illocutoire invariable comme les cas de *c'est un comble, ça suffit, il ne manquerait plus que ça, ça craint*, mais ce n'est pas toujours le cas. Néanmoins, le sens d'une formule expressive mérite parfois une observation approfondie du contexte du fait de son caractère polysémique. Nous pensons qu'une formule polysémique peut exprimer des sens qui sont liés entre eux, c'est-à-dire qu'ils se situent soit dans un noyau positif (joie-satisfaction-contentement) ou négatif (colère-indignation-mécontentement).

## 8. Conclusion

Ces observations montrent que les FEI représentent des formes « toutes faites » et font partie du stock des éléments de la langue d'une même communauté linguistique (Tutin, 2019). Ainsi, elles constituent des énoncés complets, mais pas nécessairement autonomes sur le plan syntaxique. L'emploi des FEI est plus au moins lié à des situations de production spécifique et elles ne sont interprétables qu'à l'intérieur d'une situation donnée, ce qui préfigure l'analyse en valeur pragmatique. Il s'agit des formules couramment utilisées, mais peu étudiées, notamment, sur des corpus authentiques des réseaux sociaux. Toutefois, cette question reste à décrire et mérite d'être approfondie notamment avec des corpus de tailles conséquentes. Notre corpus semble un peu réduit, mais reste assez diversifié pour une première investigation sur les FEI. Nous allons poursuivre cette étude dans le cadre de notre recherche doctorale et du projet Polonium (Pragmax) en étendant les corpus étudiés avec l'observation d'un corpus littéraire et corpus d'oral spontané.

### Références bibliographiques

- ANSCOMBRE, Jean-Claude (2011). Figement, idiomaticité et matrices lexicales, *Le figement linguistique : la parole entravée*. Paris : Honoré Champion, 17-40.
- AUSTIN, John (1970). *Quand dire, c'est faire*. Paris : Seuil.
- BALLY, Charles (1909). *Traité de stylistique française*. Paris : Klincksieck.
- BIDAUD, Françoise (2002). Structures figées de la conversation: analyse contrastive français-italien. Bern : Peter Lang.
- BLANCO, Xavier (2010). Traduction des pragmatèmes dans les guides de conversation en russe. *Contenus conceptuels et enjeux culturels, Université Autonome de Barcelone, Synergies Tunisie, 2 : 75-84.*
- BLANCO, Xavier (2013). Équivalents de traduction pour les pragmatèmes dans la Lexicographie bilingue Français-Espagnol, *Lexicographica, 29 : 5-28.*
- BLANCO, Xavier (2014). Inventaire lexicographique d'une sous-classe de phrasèmes délaissée : les pragmatèmes, *Cahiers de Lexicologie, 104 : 133-153.*
- BLANCO, Xavier (2015). Les pragmatèmes: définition, typologie et traitement lexicographique. *Verbum, 4 : 17-25.*

- BLANCO, Xavier & MEJRI, Salah (2018). *Les pragmatèmes*. Paris : Classiques Garnier.
- BURGER, Hannelore. (2010). *Phraseologie*. Berlin, Eine Einführung am Beispiel des Deutschen: Schmidt.
- DOSTIE, Gaétane (2004). *Pragmaticalisation et marqueurs discursifs. Analyse sémantique et traitement lexicographique*. Bruxelles : De Boeck, Duculot.
- DOSTIE, Gaétane (2018). Phrases préfabriquées : typologies et moules sémantiques, Séminaire Délicortal du Lidilem.
- DZIADKIEWICZ, Anne (2009). Vers un dictionnaire pragmatique français-polonais, polonais-français : quels critères pour le choix des entrées ? In *Passeurs de mots, passeurs d'espoir : lexicologie, terminologie et traduction face au défi de la diversité*. Lisbonne.
- FLECHON, Geneviève, FRASI, Paolo, POLGUERE Alain. (2012). Les pragmatèmes ont-ils un charme indéfinissable ? *Lexiques. Identités. Cultures*, 81–104.
- FONAGY, Ivan (1997). Figement et changement sémantiques. MARTINS-BALTAR M. (Éd.) *La locution entre langue et usages*, 131-164.
- GHARBI, Najwa (2018). Les pragmatèmes d'affect : délimitation définitoire et propriétés sémantico-pragmatiques, in *Lublin Studies in Modern Languages and Literature*, vol 42, N 4.
- GROSS, Gaston (1996). Les expressions figées en français. Paris : Ophrys.
- KAUFFER, Maurice (2011). Actes de langage stéréotypés en français et en allemand. Pour une redéfinition du stéréotype grâce à la phraséologie, *Nouveaux Cahiers d'allemand*, 35-53.
- KAUFFER, Maurice (2012a). Petit dictionnaire permanent des « actes de langage stéréotypés (ALS) Microstructure de dasist die Höhe ! *Nouveaux Cahiers d'allemand*, 129-145.
- KAUFFER, Maurice (2012b). Plaidoyer pour une phraséologie pragmatique sur la base des ALS français et allemands, *II International Congress of Phraseology and Paremi-ology*, Brasilia, Bresil. Pontes Editores, 2 : 192-208.
- KAUFFER, Maurice (2013a). Le figement des actes de langage stéréotypés » en français et en allemand », *Pratique*, 159-160 : 42-54.
- KAUFFER, Maurice (2013b). Tu vas voir ce que tu vas voir ! Actes de langage stéréotypés et expression de la menace. In Congrès international de linguistique et de philologie romanes (CILPR), section 5 : lexicologie, phraséologie, lexicographie, Nancy ATILF 357-368.
- KAUFFER, Maurice (2017). Les actes de langage stéréotypés. Séminaire, Délicortal, Université Grenoble Alpes & LIDILEM.
- KERBRAT-ORECCHIONI, Catherine (2005). Les actes de langage dans le discours : théorie et fonctionnement. Paris : Armand Colin.
- KLEIN, Jean René & LAMIROY, Béatrice (2011). Routines conversationnelles et figement, *Le figement linguistique : la parole entravée*. Paris : Honoré Champion, 195-214.
- LE PESANT, Denis (2015). Base de phrases figées du français. Non publié.
- LONGHI, Julien, (2013). Essai de caractérisation du tweet politique, *L'information grammaticale*, 136 : 25-32.
- LONGHI, Julien (2017). Le corpus Polititweets : enjeux institutionnels, juridiques, techniques et philologiques, Ciara Wigham et Gudrun Ledegen. *Corpus de communication médiée par les réseaux : construction, structuration, analyse*, Harmattan.
- LÓPEZ SIMÓ, Mireia (2016). Fórmulas de la conversación. Propuesta de definición y clasificación con vistas a su traducción español-francés, francés-español. Thèse de doctorat. Université d'Alicante.
- MARQUE-PUCHEU, Christiane (2007). Les énoncés liés à une situation: mode de fonctionnement et mode d'accès en langue 2, *Hieronymus*, 25-48.
- MELCUK, Igor (2004). La non-compositionnalité en morphologie linguistique, *Verbum*, 26-4 : 439-458.

- MELCUK, Igor (2008). Phraséologie dans la langue et dans le dictionnaire, *Repères & Applications* (VI), XXIVes Journées Pédagogiques sur l'Enseignement du Français en Espagne, Barcelone, 3-5 : 1-13.
- MELCUK, Igor (2011). Phrasèmes dans le dictionnaire, *Le figement linguistique : la parole entravée*. Paris : Honoré Champion, 41-61.
- MELCUK, Igor (2013). Tout ce que nous voulions savoir sur les phrasèmes, mais.... *Cahiers de Lexicologie*, 102 : 129-149.
- MERTINS-BALTAR, Michel (1997). *La locution entre langue et usages*. Fontenay Saint-Cloud : ENS Éditions.
- PANKCHURST, Rachel (2009). Short Message Service (SMS) : typologie et problématiques futures, in Arnavielle T. (coord.), *Polyphonies, pour Michelle Lanvin*, Université Paul-Valéry Montpellier 3 : 33-52.
- PAVEAU, Marie-Anne (2012). Activités langagières et technologie discursive. L'exemple de Twitter, La pensée du discours, 27/02/2012, repéré à <http://penseedudiscours.hypotheses.org/8338> (consulté le 17/05/2014).
- REY, Alain & CHANTREAU, Sophie (2005). *Dictionnaire des expressions et locutions*. Paris : Le Robert.
- SEARLE, John (1972). *Les Actes de langage*. Paris : Hermann.
- TUTIN, Agnès (à paraître). Phrases préfabriquées des interactions : quelques observations sur le corpus CLAPI.
- WEILL, Isabelle (1993). La menace comme acte de langage : étude diachronique de quelques formules de français, *Linx*, 28 : 85-105.
- ZIFONUN, Gisela, HOFFMANN, Ludger, STRCKER, Bruno (1997). *Grammatik der deutschen Sprache*. Berlin, New York: de Gruyter.

## LA CONCESSION EN FRANÇAIS DANS DES DEBATS MEDIATIQUES

Thi Thu Trang DO<sup>a</sup>

[dothithutrang12@gmail.com](mailto:dothithutrang12@gmail.com)

<sup>a</sup>Equipe d'accueil 3967 CLILLAC-ARP, Paris, France

<sup>b</sup>Laboratoire LLL, Orléans, France

### 1. Problématique et définition opératoire de la concession

La concession est un phénomène syntaxique complexe dans la mesure où il intervient au-delà du mot (même si on peut avoir, rarement, une concession entre deux mots) et qu'elle apparaît avant tout dans des contextes de subordination (on a toujours un élément subordonné par rapport à un autre élément). Mais il ne s'agit pas d'une subordination classique comme les subordinations de temps, de lieu, de cause, de conséquence, etc. Le temps, le lieu, entre autres, peuvent exister sans prise de position énonciative, mais la concession vient toujours d'un "je". Le locuteur se met toujours à la place de l'autre pour se dire : « je sais ce que tu penses, je l'admets mais j'essaie de te convaincre ». C'est pour ces raisons que la concession relève d'un processus de traitement de haut niveau et qu'il n'y a pas de définition unique du phénomène, absence corrélée à l'hétérogénéité terminologique.

Les définitions de la concession dans les dictionnaires de linguistique ont permis d'identifier trois dimensions, à savoir : linguistique, logique et interactionnelle. Une recherche bibliographique fondée sur ces trois perspectives a relevé les principes qui font consensus dans la littérature concernant les propriétés morphosyntaxiques et pragmatiques (Morel 1996, Masseron 1999), énonciatives (Paillard 1989), les perspectives d'analyse de discours (Garnier & Sitri 2008, Masseron 1999) et l'approche argumentative et conversationnelle (Moeschler & de Spengler 1982). Mais comme non seulement la terminologie de la concession n'est pas stable mais qu'en plus les définitions sont diverses, on retiendra comme définition opératoire de la concession celle qui suit :

Un énoncé à deux propositions  $p$  et  $q$  où le locuteur pose ou suppose la vérité de  $p$  ou, autrement dit, où il reconnaît la vérité ou la possibilité de la proposition  $p$  ;  $q$  est un argument contradictoire par rapport à  $p$  ; la contradiction entre  $p$  et  $q$  configure la portée de la conclusion  $r$ . Il y a généralement un élément de transition – le plus souvent un connecteur – entre  $p$  et  $q$ .

Exemple :

- (1)  $FG^1$  : *cette étude va coûter cent mille euros alors qu'on ne sait toujours pas si au final IKEA viendra (GB\_05\_03\_13)*<sup>2</sup>

$p$	$q$	$R$
cette étude va coûter cent mille euros	on ne sait toujours pas si au final IKEA viendra	il ne faut pas mener cette étude

Jusqu'à aujourd'hui, la plupart des recherches consacrées à la concession opèrent à partir d'exemples fabriqués ou de corpus écrits. Un corpus oral permettra d'identifier les phénomènes présents en français parlé. Nous avons choisi de traiter le problème dans le discours argumentatif médiatisé. Le rôle du numérique sera examiné dans la constitution du corpus (y compris le type de corpus choisi, les débats médiatisés, et les logiciels

<sup>1</sup> Initiales du nom du locuteur.

<sup>2</sup> On note GB\_05\_03\_13 le débat du Grand Bûcher diffusé le 05/03/2013.

d'enregistrement, de transcription et de présentation du corpus) et dans l'analyse de ce dernier qui a permis de relever différentes caractéristiques typiques de la concession à l'oral comme des concessions enchaînées, co-construites et des concessions définitionnelles.

## 2. Rôle du numérique dans la constitution du corpus

Ni strictement syntaxique puisqu'elle ne propose pas une liste limitée de ses moyens verbaux, ni strictement logique puisqu'elle comporte un aspect psychologique, la concession est un phénomène particulièrement pertinent pour une analyse fondée sur un corpus.

### 2.1. Constitution du corpus

Pour mener à bien ce travail, nous avons commencé par faire des requêtes dans deux grands corpus oraux en français, ESLO (<http://eslo.huma-num.fr/>) et CLAPI (<http://clapi.icar.cnrs.fr/>). Comme le premier est conçu dans une perspective sociolinguistique à partir d'enquêtes et le second contient principalement des interactions professionnelles, peu d'occurrences pertinentes pour une analyse de la concession pouvaient en être extraites. Il fallait envisager un corpus dédié.

Quel genre de discours choisir ? Le discours des médias a été privilégié car il permettait de s'affranchir des questions juridiques. Et les débats apparaissaient comme une ressource potentielle du fait de l'interactivité et des échanges argumentatifs.

La nécessité de disposer d'un corpus contenant un grand nombre d'expressions de la concession dans le français parlé contemporain, éristique et médiatique nous a conduite à choisir huit débats de l'émission *Le Grand Bûcher*<sup>3</sup> diffusée sur Radio France Bleu Orléans.

*Le Grand Bûcher* est l'émission la plus décapante du paysage radiophonique orléanais. Elle était diffusée cinq jours sur sept du lundi au vendredi, ce qui a permis un large choix des émissions les plus adaptées à notre but, c'est-à-dire celles qui présentent beaucoup de concessions.

Les sujets débattus sont très variés : événements sportifs, sociaux, culturels, économiques, politiques mais ils font tous écho à l'actualité du jour. L'émission est devenue incontournable pour beaucoup d'auditeurs du Loiret.

Les présentateurs et les polémistes ne sont pas les mêmes d'une fois sur l'autre, ce qui permet, par la diversité des intervenants, une variété dans les modalités d'expression et d'argumentation.

De plus, les invités peuvent être des experts du sujet à débattre, des fonctionnaires, des linguistes, des chefs d'entreprise, des étudiants, etc., et toujours des individualités qui sont enthousiastes à exprimer leur point de vue. Ils ne discutent pas en tant que représentants qualifiés d'un parti ou d'un syndicat (ce qui n'exclut pas leur affiliation qui est alors annoncée en début d'émission), ils sont invités à titre personnel. Leur point de vue peut coïncider avec celui du groupe auquel ils appartiennent mais ils sont libres de s'exprimer, ce qui donne aux auditeurs un regard croisé sur l'actualité dans une expression plus spontanée nous offrant quantité d'expressions argumentatives dont concessives.

Cette ressource est d'autant plus exploitable que le langage utilisé dans ces débats n'est ni appauvri ni formaté, c'est une langue orale quotidienne qui n'est pas rigide et « pauvre »

<sup>3</sup> *Le Grand Bûcher* a été diffusé sur France Bleu Orléans fréquence 100.9 tous les jours sauf le dimanche pendant une vingtaine de minutes entre 18h05 et 18h30 de septembre 2007 à juin 2013.



comme il advient souvent dans les débats politiques où les locuteurs utilisent des formules stéréotypées.

Alors que les autres chercheurs qui ont travaillé sur l'oral pour étudier la concession ont eu recours à des corpus qu'ils n'avaient pas eux-mêmes constitués (excepté le travail de Morel 1996, Masseron 1999, Vincent & Heisler 1999, Garnier & Sitri 2008), les données présentées ont été réalisées spécialement pour ce travail, en prenant en compte les propriétés orales des énoncés comme le schéma intonatif qui les structure et le contexte situationnel qui convoque tout un réseau des références sans lequel l'énoncé est privé de sa substance sémantique spécifique, comme le remarque Morel.

De plus, comme *Le Grand Bûcher* est une émission radiophonique, nous, les transcripteurs sommes des auditeurs réels et nous ne courons pas le risque de perdre les gestes, les mimiques des « bûcherons ». Huit émissions ont été retenues, soit cent soixante et onze minutes de paroles et 37.443 mots avec 124 occurrences de concession.

## 2.2. *Numérique et enregistrement du corpus*

Notre corpus se compose d'émissions radiophoniques enregistrées sur le streaming en direct avec Jaksta Streaming Media Recorder afin d'avoir des signaux numériques. Cet enregistreur est facile à utiliser pour les amateurs de multimédias, les signaux sont stables et de qualité (il n'y a pas d'interruption de signaux, ni de bruit gênant). Son originalité réside dans le fait qu'il détecte automatiquement les vidéos, musiques et radios jouées sur notre ordinateur et les télécharge. Même si nous étions absente, nous avons mis ce logiciel en marche pour faire le streaming de toutes les émissions du *Grand Bûcher* afin de ne rater aucun numéro. Une fois enregistrés, nous avons converti les fichiers dans le format WAV, compatible avec le logiciel de transcription *Transcriber*. Nous avons sauvegardé une quarantaine d'émissions pour choisir ensuite huit débats les plus animés.

## 2.3. *Numérique et transcription du corpus*

Une fois les enregistrements recueillis, ils ont été transcrits à l'aide du logiciel *Transcriber*. C'est un logiciel de transcription accessible au grand public. Il permet d'aligner le texte et le son et des sorties compatibles. Il est donc adapté à ce type de tâche. On a adopté des conventions de transcription et les politiques de dépôt et de diffusion du corpus de la BNF. Notre corpus répond ainsi à toutes les spécifications techniques et documentaires du dépôt sur les plates-formes *CoCoON* (Collection de Corpus Oraux Numériques) et *Gallica* (BnF).

Les transcriptions sont présentées sous deux formats :

- Fichier *Transcriber* (.trs) (cf. Image 1)
- Fichier *Microsoft Word* (.docx)

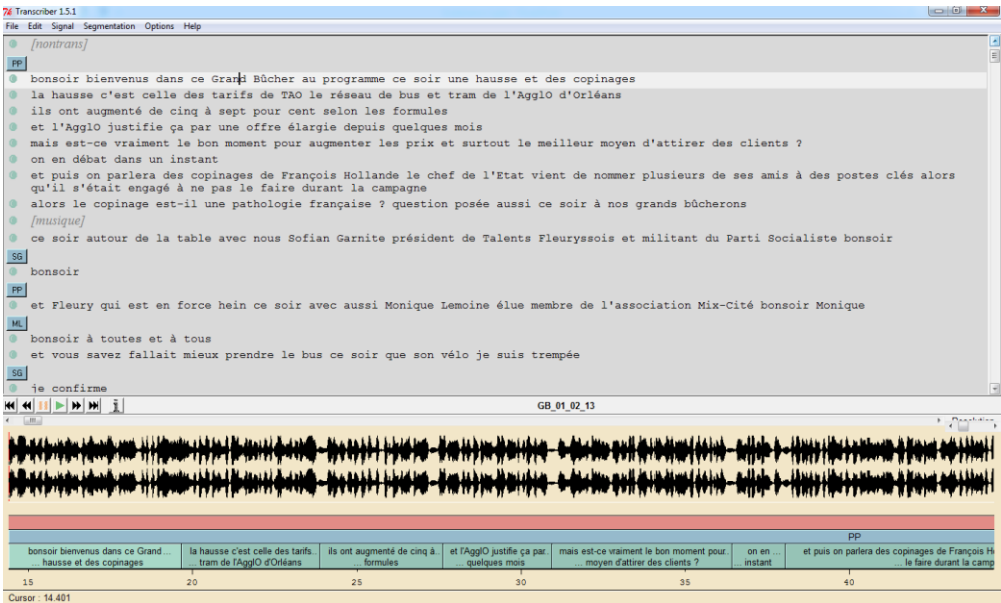


Image 11 - Capture d'écran de *Transcriber*

Grâce à l'alignement du texte et du son, on peut réécouter les segments de paroles en cours d'analyse autant de fois souhaitées et de façon rapide et simple, ce qui facilite beaucoup les analyses ultérieures.

Pourtant, on a rencontré quelques difficultés lors de la transcription du corpus.

- Au niveau de l'interaction

Ce qui pose le plus de problème dans une transcription est le chevauchement. Les locuteurs n'hésitent pas à couper la parole d'autrui dans des discussions animées. Dans certaines circonstances, il devient difficile de suivre et il n'est pas possible de conserver le détail du propos.

- Au niveau du lexique

Les locuteurs utilisent beaucoup de locutions, de tournures stéréotypées et de figements qui relèvent d'un français oral non familier pour un non-francophone surtout que le débit est beaucoup plus rapide que le nôtre. Voici quelques exemples :

*ML : je crains aussi que ça soit toujours un vieux **serpent de mer** (GB\_24\_01\_13)*

Cet énoncé avait été transcrit « je crois aussi que ça dépend des yeux de certains maires » [il s'agissait, dans la discussion, du logement social]

Ou le propos de Yves Bodard « *les choses n'arrivent pas comme un **cheveu sur la soupe*** » avait été transcrit « *les choses n'arrivent pas comme un clou à la soupe* »

- Au niveau phonétique

Comme il existe plusieurs graphies pour un phonème français, des difficultés surgissent pour la transcription. Voici un exemple :

*SB : c'est Corinne Leveleux-Teixeira elle veut piétonniser le centre-ville d'Orléans ce que l'on appelle l'**intramaïl** à l'exception de de quelques rues*

Dans cet exemple, le mot « intramail » avait été transcrit « la tramaille » – un mot qui n'existe pas.

Par ailleurs, en français, les phénomènes d'élisions, de liaisons et d'enchaînements sont très nombreux ce qui rend la transcription plus difficile.

- Au niveau du rythme et débit

Certains répètent souvent les mêmes mots déjà prononcés de sorte qu'il faut vérifier combien de fois le mot revient. En outre, comme il s'agit de langage oral, beaucoup de locuteurs ne finissent pas leurs phrases et en reprennent une nouvelle sans pause. Ces phénomènes sont aussi pour nous de grands obstacles dans la transcription.

Par exemple :

- (2) YB : et si on dév- et qu'est-ce qu'il fait ils mettent et les rares les rares sous qu'ils ont ils les dépensent pas pourquoi ? parce qu'ils sont pas sécurisés
- (3) SL : et c'est vrai que y a c'est pas parce que les comment dire les enseignants sont pour la je veux dire dans leur grande majorité ils trouvent que l'école doit comme le reste de la société évoluer et qu'ils- les réformes peuvent être [pron=peufête] peuvent être nécessaires

Face à ces difficultés, le logiciel de transcription nous a facilité la tâche en proposant au transcripteur le choix des segments, de revenir à ceux qui posent des problèmes d'écoute, de transcrire les chevauchements, ce qui a permis de mieux analyser le corpus oral.

### **3. Rôle du numérique dans l'analyse du corpus**

#### **3.1. Principe d'analyse et modélisation**

On a relevé cent vingt-trois exemples de concession dans le corpus en recourant aux divers critères avancés dans la littérature, tels le type de concession, les éléments de citation (y compris les discours rapportés ou les reprises), la symbolisation logique et les unités lexicales utilisées pour exprimer la concession (conjonctions, adverbess, prépositions, verbes, etc.). On a traité ces observables selon trois perspectives. Dans une perspective *logique*, on a examiné la concession comme un élément relevant de la causalité (au même titre que la condition, l'opposition etc.). La dimension *interactionnelle* a pris en compte les dimensions pragmatique et sociologique en interrogeant la position du locuteur dans l'échange, le contexte, les effets attendus et obtenus. Les marqueurs de concession ont fondé la partie *linguistique* de l'étude.

A la différence des concessions fabriquées ou relevées dans les textes écrits, celles apparaissant dans nos corpus oraux se signalent par un grand nombre de modalisations, de négations, de marques d'insistance, de réticence, d'évaluation, d'énonciation, de subjectivité etc. Ces éléments seront utilisés pour mener à bien les analyses.

On a adopté une démarche empirique, partant des occurrences pour tendre vers la modélisation. Des codes couleurs ont été utilisés pour marquer les catégories des moyens verbaux qui concourent à la concession en vue d'une modélisation. Par exemple, le présentatif est surligné en rose, la négation en turquoise, l'atténuation en jaune, le renforcement en rouge, etc. (cf. Tableau 1)

1	MV : c'est vrai que le la victime peut considérer que sauf que quand même il bosse il bosse dans les conditions qui peuvent être difficiles et puis le fond du problème c'est que oui il faut payer sa dette mais au-delà on peut pas la payer indéfiniment et le système carcéral français ou au moins pour le peu que je puisse en savoir il est itératif c'est-à-dire que quand on y rentre on en sort plus
2	PR-M : je sais que depuis il y a eu une concertation entre juillet et octobre donc il y a eu un certain nombre de forums de discussions se qui sont établis mais on a le sentiment que dès le départ les choses ont été en quelque sorte décidées voire même pliées

Tableau 1 - Code couleur

Des tableaux Excel permettent d'avoir une meilleure vision de la distribution des éléments dans une concession (cf. Annexe 1).

### 3.2. Quelques caractéristiques de la concession à l'oral

Le corpus oral enregistré, transcrit et analysé avec l'aide des logiciels a permis de relever quelques caractéristiques typiques de la concession en discours.

#### 3.2.1. Concessions enchaînées

On entend par *concessions enchaînées* les concessions successives de type Concession 1 – Concession 2 ou l'insertion de la Concession 1 dans la Concession 2.

Examinons un exemple :

- (4) JG : [moi je veux bien comprendre que] les gens qui nous écoutent les auditeurs quand on leur balance un chiffre comme ça de cent mille euros pour une étude ils peuvent trouver ça très cher [même moi tout à l'heure j'ai dit cent mille euros je me suis dit hors antenne tout à l'heure je me suis dit oh la vache euh je devrais me faire bureau d'étude de u-u-u- urbaniste] mais on sait pas euh ce qu'il y a dans cette étude comme ça donc voilà euh c'est sûrement justifié

Dans cet exemple, on trouve deux concessions :

Concession 1 :

$p$	Connecteur à valeur concessive	$Q$	$r$
les gens qui nous écoutent les auditeurs quand on leur balance un chiffre comme ça de cent mille euros pour une étude ils peuvent trouver ça très cher	<b>mais</b>	on sait pas euh ce qu'il y a dans cette étude comme ça	donc c'est sûrement justifié

Concession 2 :

Connecteur à valeur concessive	$p$	$q$
<b>même</b>	moi	je me suis dit oh la vache euh je devrais me faire bureau d'étude de u-u-u- urbaniste

En observant ces tableaux, on constate que la deuxième concession se trouve à l'intérieur de la première et apporte une information complémentaire en introduisant un équilibre entre *moi* et *les gens* alors que *je veux bien comprendre que* ne concourt pas vraiment à la formulation de cette concession. Pourtant, cette proposition introduit la première concession tout en anticipant ce qui sera dit concernant l'expérience du locuteur qui est actualisée par la deuxième concession.

De même, dans l'exemple 5, nous voyons deux concessions enchâssées :

- (5) NR : ça va être très difficile [parce que quand ça se passe un sur son territoire avec tous ces titres dans la musette et qu'il va falloir prouver c'est loin d'être gagné] **et en tout cas** j'espère et je souhaite même si ça ne sera pas si facile que l'équipe de France féminine fasse un bel Euro sur notre territoire en deux mille treize

Concession 1 :

<i>p</i>	Connecteur à valeur concessive	<i>q</i>
<i>ça va être très difficile</i>	<b>et en tout cas</b>	<i>j'espère et je souhaite même si ça ne sera pas si facile que l'équipe de France féminine fasse un bel Euro sur notre territoire en deux mille treize</i>

Concession 2 (qui se trouve à l'intérieur de la proposition *q* de la concession 1) :

<i>p</i>	Connecteur à valeur concessive	<i>q</i>
<i>j'espère et je souhaite que l'équipe de France féminine fasse un bel Euro sur notre territoire en deux mille treize</i>	<b>même si</b>	<i>ça ne sera pas si facile</i>

Dans ces exemples, la concession 2 se trouve à l'intérieur de la concession 1. Mais l'enchaînement des concessions peut être présenté sous forme Concession 1 – concession 2. L'exemple suivant se compose de trois concessions successives :

- (6) PR-M : je ne ferai certainement pas le procès de dire que les enseignants ne sont pas attentifs au bien-être et à l'évolution des élèves **néanmoins** je crois que comme bien souvent on essaie de faire passer les choses avant même qu'elles ne soient abouties **pourtant** le le corps enseignant est a priori assez favorable au pouvoir en place [...] ma surprise est **tout de même** de voir que l'électorat majoritairement favorable au pouvoir actuel en conteste la méthodologie

Elles peuvent se trouver dans plusieurs tours de parole, mais elles sont énoncées par un même locuteur comme dans l'exemple ci-dessous (les concessions sont notées en gras) :

- (7) ML : je voudrais rappeler quand même que Sarkozy avait essayé de mettre son fils le Prince Jean à la tête hein ? bon alors c'était carrément du népotisme  
 PP : mais c'était pas passé  
 ML : là c'était pas passé mais on se demande comment parce que ça a failli passer  
 SG : il a pas dit son dernier mot  
 ML : **or là oui voilà non mais là si vous voulez c'est quand même pas du népotisme ce sont des gens on l'espère qui sont compétents dans [pron=pi]**  
 PP : alors Olivier Schrameck il connaît rien dans l'audiovisuel quand même  
 ML : **il connaît rien il va apprendre vite**

Parmi les cent vingt-trois concessions du corpus, on en compte trente-six de type enchaîné, soit 30% environ. Ce phénomène n'a pas été relevé dans les recherches sur la concession à l'écrit, il constitue une des originalités qu'offrent les débats polémiques.

### 3.2.2. Concessions co-construites

À la différence de l'écrit, plusieurs concessions à l'oral recourent à la co-construction entre les interlocuteurs, c'est-à-dire que les idées exprimées en *p* et en *q* ne viennent pas toutes d'un même locuteur mais l'une d'entre elles vient de l'interlocuteur.

Pour être considérée comme une concession co-construite, l'idée exprimée dans une des deux propositions concessives doit se trouver dans le tour de parole qui la précède immédiatement, comme l'illustre l'exemple (8):

- (8) DT : on chasse les jeunes qui boivent sur les bords de Loire mais là on veut bien attirer un bar et #1 un bar à champagne donc #

*ML : #2 c'est pas tout à fait la même chose #*

*DT : ben c'est pas tout à fait la même chose c'est pas tout à fait la même clientèle effectivement mais bon moi je trouve que toujours ramener à la consommation*

Le nouveau cafetier, qui veut faire du lieu un bar à champagne, vise une clientèle un peu plus chic qu'auparavant. DT a repris le propos de ML pour formuler la séquence *p* de la concession mais rectifie ensuite sa portée, au contraire de ce qu'on observe dans l'exemple 9 où la reprise a lieu en *q* pour insister sur l'accord du locuteur avec son interlocuteur. C'est le seul exemple de cette catégorie.

- (9) [DT : euh moi je suis pour la gratuité des transports ça peut être aussi une façon d'emmener des gens au centre-ville de venir consommer en centre-ville plutôt que d'aller sur les extérieurs

*JH : non mais les transports c'est jamais gratuit chère madame*

*DT : les transports gratuits ça existe*

*JH : c'est payé par les impôts si par les usagers c'est les impôts]*

*DT : oui mais justement c'est déjà payé par les impôts*

La reprise peut être une reformulation comme dans l'exemple suivant :

- (10)PR-M : je je n'ai pas tout à fait terminé mon propos

*YB : eh oui mais il faut que ce soit interactif tout ça*

Pour YB, le principe du dialogue n'est pas un enchaînement de paroles mais suppose une interaction, c'est-à-dire qu'il peut y avoir des chevauchements. La co-construction de cette concession se manifeste par la réponse « oui » à l'intervention de l'interlocuteur qui peut être reformulée comme suit : « c'est vrai que vous n'avez pas terminé votre propos ».

Dans l'exemple (11), la reformulation se trouve dans la deuxième proposition de la concession :

- (11)JG : on parle toujours en France là du problème des crises de délocalisation mais on oublie de parler de tous ceux qui viennent pas et les Suédois et les autres parce que quand ils veulent arriver en France ils trouvent un les Français sont chiants deux y a des études partout #1 y a y a y a des associations qui vont faire qui vont faire une plainte contre ceci comme cela #

*BB : c'est pour ça que c'est pour ça que l'Assemblée nationale vient de voter vient de voter enfin vient d'admettre une simplification des normes administratives imposées aux collectivités parce qu'effectivement c'est déjà lourd pour les collectivités mais bien sûr comme Jerry Gras*

*le dit justement y a un certain nombre d'entreprises c'est du manque à gagner qui ne viendront jamais qui ne viendront parce que c'est trop compliqué quoi*

L'exemple (12) illustre un autre phénomène où le locuteur formule une concession en donnant suite à la proposition de son interlocuteur sans reprise ni reformulation.

(12) DV : il peut éventuellement peut-être il va y avoir une certaine décompression

PV : mais un retour de fait beaucoup plus difficile

Ici, les deux séquences de la concession sont enchaînées et sont produites par deux locuteurs. On nommera ce cas de figure concession conjointe.

On constate au total cinq possibilités de co-construction de la concession  $p - q$  :

- $p$  = reprise du propos de l'interlocuteur
- $q$  = reprise du propos de l'interlocuteur
- $p$  = reformulation du propos de l'interlocuteur
- $q$  = reformulation du propos de l'interlocuteur
- Concessions conjointes (enchaînement des idées sans reprise ni reformulation)

Les concessions co-construites se réalisent souvent pendant les chevauchements. Mais comme Transcriber permet de superposer les tours de parole, ce phénomène ne pose pas de problèmes pour la transcription et l'analyse.

Le fait de passer à la radio pourrait jouer sur le comportement verbal des locuteurs : ils aimeraient prendre la parole pour marquer leur présence mais atténuent en même temps leurs propos en donnant un accord partiel à l'interlocuteur pour former des concessions co-construites.

### 3.2.3. Concessions définitionnelles

Un des résultats de l'analyse du corpus oral concerne la caractérisation des « concessions définitionnelles » dont voici un exemple :

(13) c'est vrai que le la victime peut considérer que sauf que quand même il bosse il bosse dans les conditions qui peuvent être difficiles et puis le fond du problème c'est que oui il faut payer sa dette mais au-delà on peut pas la payer indéfiniment et le système carcéral français ou au moins pour le peu que je puisse en savoir il est itératif c'est-à-dire que quand on y rentre on en sort plus

Le locuteur commence par ratifier une définition de la justice comme paiement d'une dette dans le système carcéral, mais il reformule cette affirmation en avançant qu'il doit y avoir extinction de l'obligation.

La concession définitionnelle se décomposerait, quand elle exploite l'ensemble des ressources de structuration, en cinq éléments : un présentatif (PRES), une modalisation (MOD), un terme générique ou abstrait (TERME), une négation (NEG) et un connecteur à valeur concessive (CC), soit :

(14) [je trouve que (MOD)] l'un dans l'autre on- on parle- on parle de cette assurance chômage de cette indemnisation de chômage en termes de [charge (TERME)] [mais (CC)] [je pense aussi que (MOD)] la société [peut (MOD)] s'y retrouver dans la mesure où elle n'abandonne [pas (NEG)] [complètement (MOD)] les gens qui sont demandeurs d'emploi parce que [je pense que (MOD)] de toute façon [il y aura (PRES)] un effet boomerang beaucoup plus redoutable

Ce type de concession représente le tiers des concessions relevées.

Outre le connecteur à valeur concessive (en faisant abstraction des constructions parataxiques), comment interpréter la présence de chacun des composants ?

- la redéfinition nécessite un *présentatif* qui opère le décrochage métalinguistique que la *modalisation* actualise en tant qu'alternative décidée par le locuteur,
- l'*hyperonyme* permet de circonscrire l'espace définitionnel réunissant deux séquences en opposition qui font intervenir la *négation*.

Comme ces débats radiophoniques abordent des sujets d'actualité, il est souvent nécessaire que les locuteurs à un moment ou un autre reviennent à la définition du terme pour clarifier certaines notions mais prennent en même temps un recul face à la compréhension usuelle de ce terme. Cette catégorie représente quarante concessions sur cent vingt-quatre, soit un tiers de notre corpus.

#### 4. En conclusion

Le numérique joue un rôle important dans la constitution du corpus. Le débat semble un choix sûr parce que c'est en débattant qu'on argumente le plus. Le débat médiatisé présente un autre avantage : on ne court pas le risque de violer le droit d'auteur.

L'analyse du corpus oral des émissions radiophoniques en streaming a montré que la concession ne peut être réductible à une liste des conjonctions car les plus souvent mentionnées dans les grammaires comme *bien que, quoique,...* sont rares, voire absentes. Ce manque de marqueurs syntaxiques est comblé par des moyens verbaux et non-verbaux comme l'intonation, la prosodie, etc.

Parmi cent vingt-trois concessions, quarante qui portent essentiellement sur la valeur des mots (comme une renégociation de leur définition) se composent d'unités lexicales hétérogènes successives (présentatif, nom générique, modalisateurs, négation...) cadrant la séquence en question et en signalant la valeur.

L'utilisation du logiciel de transcription *Transcriber* a permis de superposer les tours de parole chevauchés en toute simplicité et de relever des occurrences de concessions enchaînées, conjointes ou co-construites assez nombreuses dans notre corpus.

Les débats médiatisés ont offert des occurrences de concessions avec des caractéristiques particulières qui n'apparaissent pas dans les corpus écrits.

Sur la base d'un script, on a calculé le temps de paroles de chaque locuteur pour voir s'il y a une corrélation entre le temps de paroles et la fréquence de production des concessions (cf. Tableau 2).

Enregistrements	Locuteurs	Temps en seconde	Millisecondes	Temps en heure	Pourcentage
GB_01_02_13	PP	331	596	0:5:31.596	23.59866975
	SG	313	436	0:5:13.436	22.30627828
	ML	419	168	0:6:59.168	29.83090025
	SD	239	948	0:3:59.948	17.07636283

**Tableau 2 - Temps de paroles des locuteurs dans l'enregistrement GB\_01\_02\_13**

Pour l'instant, cet indice n'est pas significatif. On se demande si la disproportion des occurrences de concession entre les locuteurs peut avoir un lien avec leur catégorie socio-professionnelle, leur tranche d'âge, etc.



Cette recherche présente une perspective qui exploite fortement la prosodie et les différents registres expressifs de la concession. Elle permet ainsi d'envisager conjointement les aspects lexico-sémantiques, syntaxiques, pragmatiques et interactionnels de l'argumentation.

### **Références bibliographiques**

- DO, Thi Thu Trang (2016). L'expression de la concession en français parlé : Le cas de la concession comme redéfinition. Colloque CEDIL'14, Grenoble, <https://hal.archives-ouvertes.fr/hal-01252383>, (nd). Consulté en mai 2018.
- GARNIER, Sylvie, SITRI, Frédérique *et al.* (2008). Problèmes épistémologiques en analyse du discours : le cas de l'énoncé concessif, in *Homenagem a Michel Pêcheux. 25 anos de presença na análise do discurso*, Campinas : SP Mercado de Letras, 99-106.
- LE PESANT, Denis (2005). Causalité et concession. *Questions de classification en linguistique : méthodes et description. Mélanges offerts à Christian Molinier*, 78 : 195-210.
- LE PESANT, Denis (2006). De la concession à la cause, et de la cause à la condition. *La cause : approche pluridisciplinaire, Linx*, 54: 61-71.
- LÉTOUBLON, Françoise (1983). *Pourtant, cependant, quoique, bien que* : dérivation des expressions de l'opposition et de la concession, in *Connecteurs pragmatiques et structures du discours, Cahiers de Linguistique Française*, 5: 85-110.
- MARTIN, Robert (1987). *Langage et croyance. Les « univers de croyance » dans la théorie sémantique*. Liège : Mardaga.
- MASSERON, Caroline (1999). *La concession de la phrase au discours. Aspects linguistiques et didactiques*. Thèse de doctorat en sciences du langage, Université Paul Verlaine-Metz.
- MOESCHLER, Jacques & SPENGLER, Nina (de) (1981). *Quand même* : de la concession à la réfutation, *Cahiers de linguistique française*, 2: 93-112.
- MOESCHLER, Jacques & SPENGLER, Nina (de) (1982). La concession ou la réfutation interdite. Approches argumentative et conversationnelle in *Concession et consécution dans le discours, Cahiers de Linguistique Française*, 4: 7-36.
- MOREL, Mary-Annick (1996). *La concession en français*. Paris : Editions Ophrys.
- PAILLARD, Michel (1989). Les chemins de la concession : quelques contrastes entre anglais et français, *Travaux linguistiques du Cerlco*, 6: 207-226.
- SOUTET, Olivier (2008). Des concessives extensionnelles aux concessives simples. Contribution à l'étude de la genèse sémantique et historique des locutions conjonctives concessives du français. *Linx*, 59: 115-132.
- VINCENT, Diane & HEISLER, Troy (1999). L'anticipation d'objections : prolepse, concession et réfutation dans la langue spontanée. *Revue québécoise de linguistique*, 27(1): 15-31.

## Annexe

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1		Présentatifs				Termes génériques ou abstraits	Négation	Modalisation								
2		classique		discours				Renforcement				Atténuation				
3		définitionnel	existentiel	définitionnel	existentiel			V. modaux	Adv	Loc.	V. d'opinion	V. modaux	Adv	Loc	V d'opinion	
4	PR-M: c'était pas exactement ma spécialité mais j'ai quand même un avis personnel	1				1	1		1							
5	SG: ça correspond au centre-ville historique mais je veux dire on n'est pas on va pas jusqu'à place d'Arc par exemple y a c'est quand même délimité	1		1			1									
6	je suis pas du tout expert sur la question euh de de de mobilité simplement en tant qu'usager ce que je remarque c'est que y a quand même une perte de qualité depuis la seconde ligne du tram		1	1		1	1		1				1		1	
7	PP: Monique Lemoine on va dire que c'est jamais le bon moment de toute façon pour augmenter des tarifs mais là quand même ça tombe comme l'a dit Sophian au moment où on a refait le réseau où y a eu des difficultés	1	1	1		1							1		1	

Annexe 1 - Distribution des éléments de la concession définitionnelle

## **LE NUMERIQUE DANS LES PRATIQUES LANGAGIERES**

# IMPACT DE LA TRANSITION NUMERIQUE SUR LA PRODUCTION DE RESSENTIS

Pauline SOUTRENON<sup>a, b</sup>

[psoutrenon@comongo.fr](mailto:psoutrenon@comongo.fr)

<sup>a</sup>SAS Comongo, Grenoble, France

<sup>b</sup>UMR Litt&Arts (5316), Grenoble, France

## 1. Introduction

Dans le cadre d'une thèse CIFRE avec la startup Comongo, nous présentons un travail de recherche en sciences du langage et humanités numériques sur l'analyse automatique de production de ressentis. Nous avons deux objectifs : définir et évaluer l'image de marque d'un client et l'aider à anticiper les ressentis et les attentes de communication de publics cibles en vue d'améliorer cette image de marque. Dans notre démarche, nous employons le terme d'analyse de ressentis qui se distingue de l'analyse de sentiments ou d'opinions. Nous travaillons sur des productions écrites (des réponses à des questions ouvertes) qui nous permettent de comparer une identité souhaitée d'un client avec une image perçue. Nous cherchons à extraire les ressentis et à effectuer une représentation synthétique de ceux-ci.

Pour cela, nous bénéficions d'un cadre industriel au sein de la startup Comongo qui souhaite effectuer une transition numérique. Un des aspects qui nous intéresse particulièrement est de prendre en compte et d'analyser, dans les pratiques de Comongo, l'impact de la transition numérique sur la production de ressentis par des groupes de personnes.

Nous posons l'hypothèse que la transition à une pratique numérique, en présentiel comme en distanciel, modifie les contenus rédigés par les participants. Identifier et caractériser ces changements est par conséquent essentiel au bon fonctionnement de l'entreprise et au bon usage des outils de traitement de langues qui accompagnent l'analyse humaine des contenus.

## 2. Constitution des données

La startup Comongo a mis en place une méthode permettant d'évaluer et d'améliorer l'image de marque d'une personne physique ou morale. Cette méthode se déroule en quatre étapes comme le montre la Figure 1 ci-dessous. Tout d'abord, un sujet est défini par un client et une identité souhaitée est éventuellement constituée avec celui-ci. Dans un deuxième temps, le client définit un panel de publics cibles à interroger. Puis, les réponses sont recueillies via une interface numérique et ces réponses constituent l'image perçue (étape 3). Enfin, la dernière étape est la comparaison entre l'identité souhaitée et l'image perçue. Cette comparaison de texte fait ressortir les lignes convergentes et divergentes pour ainsi émettre des préconisations en termes d'axes de communication avec un accompagnement humain de la démarche du client dans l'amélioration et l'évaluation de son image de marque.

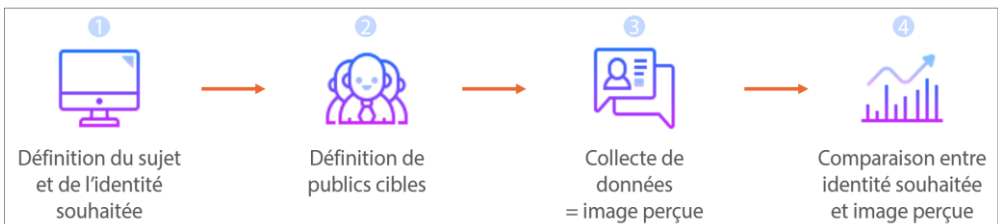


Figure 12 - La méthode de Comongo

Les données recueillies lors de la troisième étape sont des réponses à des questions orientées de différents types :

- Les évocations : ce que le sujet évoque ;
- Les polarités positives et négatives : ce qui est apprécié ou non vis-à-vis du sujet ;
- L'unicité : ce qui rend unique le sujet ;
- Les attentes et les manques : ce qui est attendu.

### 2.1. Formes d'énoncés

Lorsqu'ils répondent sur l'interface numérique, nous demandons aux audités de remplir une idée par champ. Une idée est une unité d'information telle que conçue par les personnes remplissant le questionnaire, ce qui implique une variabilité de la forme et de la structure. Des exemples sont donnés dans le Tableau 1 ci-dessous, il s'agit de réponses au sujet d'une ville. Chaque cellule correspond à une idée qui a été donnée par une personne. Les données recueillies peuvent être de formes différentes : nous pouvons trouver des mots, des groupes de mots ou encore des phrases.

Mots	Groupes de mots	Phrases
Montagne	Pistes cyclables	Une obligation de travail mais pas une envie de rester y vivre
Sport	Une ville à taille humaine	Il manque des entreprises de taille moyenne
Dynamisme	Un espace de rencontre	
Pollution	La proximité de la montagne	
Chaleur	Un cadre de vie idéal pour les sportifs	

Tableau 9 - Les données recueillies

### 2.2. Volumétrie

Nous constatons que la volumétrie est variable selon les études. En effet, selon la personne répondant au questionnaire et ses motivations vis-à-vis du sujet, elle peut être plus ou moins productive en termes de nombre d'idées.

Le Tableau 2 ci-dessous montre la volumétrie des études. Nous pouvons constater que les études réalisées en présentiel numérique (études  $\beta$ ,  $\gamma$  et  $\delta$ ) ont tendance à générer un nombre moyen d'idées par audité plus faible que l'étude  $\alpha$  réalisée en présentiel papier.

Etude	Nombre d'audités	Nombre d'idées	Nombre moyen d'idées par audité	Pratique
Etude $\alpha$	29	693	23,9	Présentiel papier
Etude $\beta$	17	219	12,9	Présentiel numérique
Etude $\gamma$	13	260	20,0	Présentiel numérique
Etude $\delta$	28	516	18,4	Présentiel numérique

Tableau 10 - Volumétrie des études

Quelque soit le nombre d'idées par étude, le but est de savoir si les mêmes idées sont répétées chez les différents audités. Pour cela, nous souhaitons automatiser partiellement le processus d'analyse en utilisant des ressources fondées sur l'intelligence humaine et permettant de repérer :

- Les mots dans les énoncés ayant des lemmes communs avec le Lexique des Formes Fléchies du Français (Clément et *al.*, 2004)
- Les idées qui appartiennent au même concept sémantique ou à concept similaire avec Diko (Lafourcade, 2007)
- La polarité, afin d'identifier si une idée est plutôt positive ou négative : LikeIt (Lafourcade et *al.*, 2015) et French Expanded Emotion Lexicon (Abdaoui et *al.*, 2016)

Cette analyse automatique est ensuite vérifiée et corrigée si besoin par le consultant. Il s'agit d'une approche incrémentale des ressources (voir Figure 2 ci-dessous) : les ressources

permettent l'annotation du corpus, chaque idée est associée à un concept. Une fois le corpus analysé, il nous sert à améliorer les ressources. Cet enrichissement est réalisé par le consultant qui peut corriger l'annotation. Ainsi, chaque nouvelle étude est enrichie de la précédente.

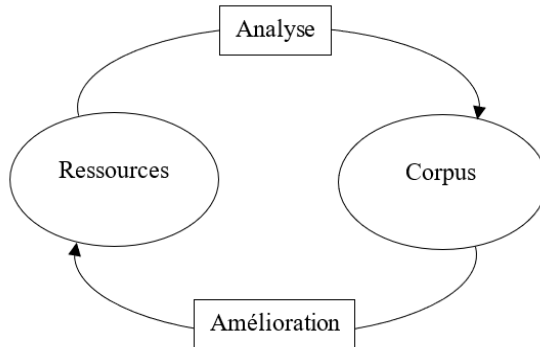


Figure 13 - Enrichissement des ressources

### 3. Transition numérique et distancielle

Nous sommes partis d'une pratique traditionnelle de Comongo d'évaluation et d'amélioration d'image de marque en présentiel papier. Le consultant Comongo aide le client dans la rédaction de son identité souhaitée et il accompagne les audités lors d'entretiens individuels ou en *focus groups* (petits groupes de 6 à 10 personnes) pour établir un diagnostic d'image perçue. Pendant ces entretiens, le consultant recueille des données qualitatives écrites sur papier par les audités. La comparaison entre l'identité souhaitée et l'image perçue est ensuite effectuée manuellement par le consultant.

Afin d'effectuer la transition numérique visée par l'entreprise, où l'utilisateur se retrouvera seul face à l'outil (pratique distancielle numérique), nous passons par une pratique transitionnelle en présentiel numérique. Au lieu de répondre sur papier, les audités saisissent leurs réponses directement sur l'interface numérique de recueil d'opinions et le consultant est présent pour accompagner la démarche. Les réponses peuvent être revues et modifiées si besoin par les audités avant de les envoyer, ce qui permet de développer chaque idée plus facilement que sur papier.

La mise en place de la pratique transitionnelle nous permet de comparer deux états de la donnée (présentiel papier et présentiel numérique) et d'identifier les impacts du numérique sur la méthode et les données. Lors du passage en distanciel, nous évaluerons les changements dus à l'absence du consultant (passage en distanciel). La comparaison des trois pratiques est présentée sur la Figure 3 ci-dessous. Les changements non exhaustifs entre les pratiques ont été mis en couleurs.



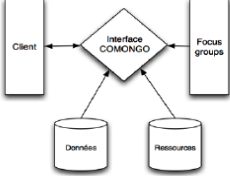
Pratique actuelle : présentiel papier	Pratique transitionnelle : présentiel numérique	Pratique voulue : distanciel numérique
		
<b>Etat initial</b>	<b>Etat transitionnel : numérique</b>	<b>Etat final : distanciel individuel</b>
Réponses personnelles Papier non conservé	Réponses personnelles Données conservées Données anonymisées Données analysables	
Réponses libres sur papier	Réponses sur l'interface numérique	
Cadre contraint : - Pression du consultant - Effet de groupe en <i>focus group</i> - Temps et moment contraints		Cadre non contraint : - Pas de pression du consultant - Pas d'effet de groupe (individuel) - Temps et moment non contraints (pas de contrôle)

Figure 14 - Comparaison des trois pratiques

Les deux changements majeurs de la pratique actuelle à la pratique transitionnelle apparaissent en vert. Premièrement, le changement de format par le biais de la plateforme numérique risque d'avoir un impact sur la qualité des réponses. La lecture demande de l'attention (Lachaux, 2011) et, sur la plateforme, les audités sont amenés à lire un tutoriel et les questions auxquelles ils doivent répondre. Carr (2012) évoque un problème de lecture avec l'utilisation d'Internet : plus Internet est utilisé et plus il est difficile de se concentrer sur une seule chose et d'être attentifs sur de longs textes. L'attention est une ressource rare, cela est notamment dû à la surabondance informationnelle à laquelle nous sommes confrontés chaque jour (Citton, 2014). Il existe une quantité importante de questionnaires en ligne, les audités seront donc peut-être moins attentifs et cela aura un impact sur la donnée qui risque d'être moins qualitative. De plus, les usagers de la plateforme n'auront pas le même rapport aux outils numériques, il faut donc aussi considérer la question de la littératie numérique. Selon l'OCDE (2000), la littératie est :

« l'aptitude à comprendre et à utiliser l'information écrite dans la vie courante, à la maison, au travail et dans la collectivité en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités ».

La littératie numérique peut donc être définie comme la compétence de chaque usager à utiliser un outil numérique.

Deuxièmement, le fait que les réponses soient enregistrées dans une base de données les inscrit de manière plus ou moins pérenne dans un système informatique. Il ne s'agit donc plus d'une donnée volatile, ce changement de mode de communication peut être la cause de freins psychologiques chez certains usagers. A contrario, le « bouclier » que représente l'écran peut permettre à d'autres de s'exprimer plus librement, sans crainte d'être jugé par le groupe.

En rose, dans la Figure 14 ci-dessus, apparaissent les changements majeurs de la pratique transitionnelle vers la pratique voulue, à savoir : le cadre de la production qui n'est plus contraint, il n'y aura plus de pression du consultant ni d'effet de groupe et nous n'aurons plus

de contrôle sur les conditions dans lesquelles les audités répondront au questionnaire (le temps passé à répondre et le moment pendant lequel ils répondent).

#### 4. Impact du numérique

Les premières observations de notre corpus par rapport à l'impact du numérique montrent qu'il existe des variations dues au cadre technologique.

La transition numérique pouvant avoir un impact sur les réponses des audités, une des hypothèses de départ était que nous risquions d'avoir moins de données en numérique que sur papier. En effet, dans la pratique transitionnelle de Comongo les audités notent leurs réponses directement dans l'outil qui les enregistre dans une base de données. Cela peut constituer un frein psychologique pour les audités qui pourraient s'autocensurer.

Cependant, nous avons constaté que le fait de répondre sur une interface numérique sans le biais du consultant rend l'expression plus libre. Bien que l'étude  $\alpha$  réalisée en présentiel papier comporte le plus grand nombre d'idées et de tokens par rapport aux trois études réalisées en présentiel numérique (section 1.2). Nous avons remarqué que les audités argumentent plus sur l'interface numérique. Ils ont tendance pour chaque idée à faire des phrases plus longues en termes de nombre de caractères et de nombre de tokens (Tableau 3 ci-dessous) : nous avons pris les productions les plus longues (nombre de caractères) de l'étude  $\alpha$  en présentiel papier et de l'étude  $\gamma$  en présentiel numérique.

Etude	Pratique	Nombre de caractères	Nombre de tokens	Production
Etude $\alpha$	Présentiel papier	130	25	« Me fournir des alertes sur les sites web dont le business model est d'aller chercher des données et de les exploiter à mes dépends »
Etude $\gamma$	Présentiel numérique	285	48	« Un avis trop tranché parfois : sur certains projets ou nous avons des impératifs ou choisis tel ou tel format, il faudrait que X comprenne que nous devons avancer avec ce support et pas proposer absolument autre chose (se recentrer davantage sur le besoin client et mieux l'écouter) »

Tableau 11 - Comparaison du nombre de caractères

L'exemple de production dans le Tableau 4 ci-dessous est une réponse d'un audité à une question de type rejet (ce qui n'est pas apprécié vis-à-vis du sujet) pour laquelle il est demandé de renseigner une idée par champ dans l'interface (chaque champ est représenté dans le tableau par une cellule). La première idée répond à la question posée, puis l'audité complète et va plus loin dans la réflexion en donnant « quelques pistes d'amélioration ». Les audités précisent davantage leur pensée sur l'interface que sur papier.

Production
« Rien que je n'apprécie pas
<b>En revanche</b> , quelques pistes d'amélioration... sur deux points essentiels
Premier point : mieux définir les critères d'une stratégie de GO/NO GO sur certains projets. XXX n'a peut-être pas vocation à accepter toutes les missions, qui risqueraient de désorganiser/perturber/mettre trop de pression sur son équipe
Deuxième point : peut-être passer plus de temps à mieux cerner le besoin client pour être sûrs de bien définir une réponse adaptée, chiffrée au plus près »

Tableau 12 - Exemple de production

#### 5. Conclusions et perspectives

Les outils du Traitement Automatique des Langues permettent d'accompagner et de faciliter les tâches du consultant, que ce soit pour le recueil des données, l'analyse de celles-ci ou leur visualisation. Notre système est fondé sur une approche centrée utilisateur : la machine



accompagne dans la réflexion et donne d'autres perspectives sur l'information, le consultant garde la maîtrise de l'interprétation et sous-traite à la machine l'organisation de l'information. La transition numérique modifie les contenus rédigés par les audités. Nous avons constaté une baisse du nombre d'idées entre l'étude  $\alpha$  réalisée en présentiel papier et les études réalisées en présentiel numérique. Il est prévu de réaliser d'autres études papier une fois la pratique distancielle numérique mise en place. Cela nous permettra de vérifier nos hypothèses et d'avoir une meilleure représentativité. Nous avons également remarqué que les réponses recueillies sur l'interface sont plus longues et plus complètes que celles recueillies sur papier en termes de nombre de caractères et de tokens. Dans les travaux à venir, il serait intéressant de mener une analyse génétique sur les études en traçant les ratures, les déplacements ou les insertions de texte en numérique et en papier.

Actuellement, nous commençons à nous diriger vers le distancielle numérique et nous avons déjà remarqué que les audités préfèrent répondre sans la présence du consultant. Nous nous apercevons également que sans la contrainte temps en distancielle, les audités ne vont pas au bout de leur pensée, mais reviennent dans l'outil plus tard pour compléter, ce qui enlève de la spontanéité.

Les altérations des contenus existent, qu'elles soient dues à la transition numérique ou à la transition de la modalité présenteielle à distancielle, il est essentiel d'en tenir compte dans la pratique d'entreprise mais aussi dans l'usage même des corpus. Il est primordial d'identifier les conditions de productions qui vont faire varier les ressources nécessaires aux outils.

### **Références bibliographiques**

- ABDAOUI, Amine, AZE, Jérôme, BRINGAY, Sandra & PONCELET, Pascal (2016). FEEL : a French Expanded Emotion Lexicon. *Language Resources and Evaluation*, (pp. 1–23).
- CARR, Nicolas (2012). *Internet nous rend-il bête ? Réapprendre à lire et à penser dans un monde fragmenté*. Paris : Robert Laffont.
- CLÉMENT, Lionel, LANG, Bernard, & SAGOT, Benoît (2004). Morphology based automatic acquisition of large-coverage lexica. In *LREC 04* (pp. 1841–1844). Lisbonne, Portugal.
- CITTON, Yves (2014). *Introduction à L'économie de l'attention. Nouvel horizon du capitalisme ?* Paris : La découverte.
- LACHAUX, Jean-Philippe (2011). *Le Cerveau attentif*. Paris : Odile Jacob.
- LAFOURCADE, Mathieu (2007). Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *SNLP'07 : 7th International Symposium on Natural Language Processing* (pp.7). Pattaya, Chonburi, Thailand.
- LAFOURCADE, Mathieu, LE BRUN, Nathalie & JOUBERT, Alain (2015). Vous aimez ?...ou pas ? LikeIt, un jeu pour construire une ressource lexicale de polarité. In *TALN : Traitement Automatique des Langues Naturelles*. Caen, France.
- OCDE (2000). *La littératie à l'ère du numérique*. Technical report, Organisation de coopération et de Développement Numérique.

## NEOLOGIE LEXICALE ET MEDIAS SOCIAUX : LE CAS DES NEOLOGISMES PAR ABREGELEMENT DE L'ANGLAIS

Pauline LEFEBVRE<sup>a</sup>

[pauline.lefebvre@univ-poitiers.fr](mailto:pauline.lefebvre@univ-poitiers.fr)

<sup>a</sup>Laboratoire FoReLLIS (EA 3816), Maison des sciences de l'homme et de la société, Poitiers, France

Ces dernières années, les médias sociaux, en tant que terrain de recherche, suscitent un intérêt grandissant. Pour le linguiste, les médias sociaux offrent l'accès à une base de données langagières massive. Ces données langagières sont constituées de publications et d'échanges entre internautes pouvant être soit réfléchis et travaillés soit, à l'inverse, très spontanés. C'est précisément parce que ce type de média offre la possibilité aux utilisateurs de s'exprimer spontanément que nous avons choisi les médias sociaux comme terrain d'exploration à une étude sur les néologismes de l'anglais. Nous pensons ainsi avoir accès à des formations néologiques créées plus instinctivement par des utilisateurs que dans un contexte plus normatif. Cette étude s'inscrit dans un projet de thèse plus large qui entreprend d'étudier les néologismes par abrègement de l'anglais sous un angle lexicographique, morphologique et phonologique. L'objectif de cette étude particulière sera de voir si l'emploi de néologismes par abrègement en anglais dans les médias sociaux est uniquement motivé par l'économie linguistique qui désigne la tendance pour un individu à réduire son effort langagier à ce qui est nécessaire à la transmission d'un message.

Après avoir défini les concepts-clés de notre étude et présenté nos hypothèses de travail, nous détaillerons comment nous avons constitué le corpus de néologismes sur lequel repose cette étude. Nous formulerons ensuite les observations que nous avons notées à partir de l'analyse du corpus.

### 1. Définitions

#### 1.1. Les médias sociaux

Avant d'aller plus loin, il convient de revenir sur ce que nous entendons ici par *médias sociaux*. Nous nous appuyerons sur la définition proposée par Kaplan et Haenlein (2010) :

In our view – and as used herein – Social Media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content. (Kaplan & Haenlein, 2010 : 61)

Les médias sociaux peuvent donc être définis comme construits à partir du Web 2.0, mais permettant la publication de contenu généré par les utilisateurs, c'est-à-dire le contenu créé et publié directement par les utilisateurs finaux, par opposition au contenu créé par les entreprises de médias traditionnelles (Kaplan & Haenlein, 2009). Pour cette étude, nous avons donc entrepris de construire un corpus de néologismes à partir de médias aussi variés que les réseaux sociaux, les blogs, les forums de discussion et les jeux vidéo en ligne.

#### 1.2. Néologie lexicale et néologismes

Si l'on se penche sur la littérature scientifique, on ne peut que constater la difficulté à définir avec précision ce qu'est un néologisme. On est alors tenté de regarder du côté de l'étymologie du mot : *neos* (nouveau), *logos* (parole, discours). Un néologisme serait donc simplement un mot nouveau. On constate assez vite les limites d'une telle définition lorsqu'on entreprend de

décider si un mot est à considérer comme néologisme ou s'il a déjà été lexicalisé, c'est-à-dire s'il a intégré le lexique commun. C'est le constat fait dans une étude réalisée par l'équipe de l'ERA 353 pour le numéro 36 de la revue *Langages* (Gardin, Lefèvre, Marcellesi & Mortureux, 1974). Au cours de cette étude dont la consigne était « relevez les unités qui vous paraissent néologiques » à partir de huit pages de l'hebdomadaire *Le Point*, il a été observé une grande disparité quant au degré d'accord manifesté par les membres de l'équipe sur le caractère néologique d'une unité (Gadin et al., 1974 :36). Ainsi, des mots apparus récemment dans la langue qui pourraient sembler nouveau à certains ne le seront pas nécessairement pour d'autres.

On ne peut pas non plus se fier au dictionnaire pour savoir si un mot a déjà intégré le lexique commun ou s'il est toujours néologisme, comme le font remarquer Pruvost et Sablayrolles (2019) :

Un mot, une expression ne sont pas non néologiques parce qu'ils sont dans le dictionnaire, mais c'est l'inverse qui est vrai : c'est parce qu'ils ne sont plus néologiques qu'ils sont intégrés dans le dictionnaire. (Pruvost & Sablayrolles, 2019 : 60)

Nous ajouterons à cela que certains dictionnaires ont justement pour vocation de compiler des néologismes, comme c'est le cas du dictionnaire en ligne *Urban Dictionary* dans lequel n'importe quel utilisateur peut proposer une définition (parfois d'ailleurs à but humoristique) pour un mot qu'il aurait entendu, lu, employé ou bien qu'il aurait créé. La présence d'un mot dans ce type de dictionnaire ne saurait donc attester formellement de la lexicalisation de ce mot.

Pour cette étude, nous définirons donc le néologisme comme un mot nouveau ou apparu récemment dans la langue. Si cette définition ne permet toujours pas de cerner ce qu'est un néologisme, elle permet au moins de tenir compte des problématiques que nous venons de mentionner.

### 1.3. *Les néologismes par abrègement*

Comme nous l'avons mentionné en amont, ce travail porte spécifiquement sur les néologismes par abrègement en anglais, c'est-à-dire les néologismes créés au moyen de la réduction d'un signifiant. Les néologismes par abrègement sont donc des néologismes de forme, par opposition aux néologismes de sens qui consistent en l'attribution d'un sens nouveau à une forme déjà existante. Trois procédés de formation lexicale nous intéressent particulièrement. Nous étudierons la *truncation*, qui consiste en la suppression d'une ou plusieurs syllabes d'un mot source (*vegetable* /'vedʒtəbəl/ → *veg* /vedʒ/) et qui est, en premier lieu, un phénomène oral mais que l'on retrouve à l'écrit en contexte informel. La truncation ne doit pas être confondue avec l'abréviation qui est un phénomène que l'on ne trouve qu'à l'écrit et consiste en la réduction d'un mot sans se soucier de pouvoir le prononcer à l'oral. Nous étudierons également l'*acronymie* est un procédé qui ne retient que les initiales d'une séquence de mots pour former un nouveau mot (*Fear Of Missing Out* → *FOMO*<sup>1</sup> /'fəʊməʊs/). Au cours du XX<sup>ème</sup> siècle, la création d'acronymes fut très populaire dans les milieux militaires avant de s'étendre aux milieux médical, scientifique et informatique (Fischer, 1998). Il convient de veiller à bien distinguer les acronymes, dont l'ensemble des initiales sera lu comme un seul mot, des sigles, dont chaque initiale sera épelée (par exemple *FYI* /'efwaɪ'aɪ/, pour *For Your Information*). Enfin, les *amalgames* (aussi appelés *mots-valise*) sont créés à partir de deux (occasionnellement trois) mots tronqués et/ou assemblés pour n'en former qu'un seul

<sup>1</sup> *FOMO* se définit comme la peur de manquer quelque chose d'intéressant auquel d'autres prennent part (*Merriam-Webster*, en ligne).

(*information + commercial* → *infomercial*<sup>2</sup>). C'est un procédé que l'on rencontre fréquemment dans la presse et dans la publicité.

Parce qu'ils consistent en la réduction d'un signifiant, ces trois procédés de formation de mots s'inscrivent dans une logique d'économie linguistique décrite ainsi par Tournier (1991) :

Lorsqu'un usager de la langue produit un énoncé et transmet ainsi une information, il le fait au prix d'un certain effort. L'effort que cela lui coûte – et ce qu'on appelle le coût linguistique – a deux aspects. Il a un aspect physique (articulatoire pour l'énoncé oral, plus généralement musculaire pour l'énoncé écrit) et un aspect mental, qui correspond à l'effort de mémoire [...] Ainsi, tout énoncé implique une transmission d'information et un coût. On appelle économie linguistique le rapport de l'information au coût. Ceci revient à dire que pour une même quantité d'information, plus le coût est grand, plus l'économie est faible. Or le coût est en rapport direct avec la longueur des lexies : plus une lexie utilisée est courte, plus l'effort musculaire, et sans doute mémoriel, est réduit. [...] Par ailleurs, le coût est en rapport inverse avec la fréquence d'emploi : plus un mot est fréquent, moins son utilisation nécessite d'effort mémoriel, indépendamment de sa longueur. (Tournier, 1991 : 63)

Selon cette logique, en réduisant le signifiant nécessaire à la transmission d'une idée, l'emploi de néologismes par abrègement serait motivé, du moins en partie, par cette tendance à la réduction du coût linguistique. Nous proposons donc d'explorer ce lien entre économie linguistique et l'emploi de ce type de néologismes spécifiquement dans les médias sociaux.

## 2. Hypothèses de travail

La communication via les médias sociaux s'organise autour de plusieurs contraintes dont certaines qui pourraient laisser penser qu'elles favoriseraient encore davantage cette tendance à l'économie linguistique. Crystal (2001) évoquait déjà le fait que la taille des écrans par lesquels nous interagissons sur Internet avait tendance à diminuer (notamment du fait que l'on accède de plus en plus à Internet via un smartphone). L'utilisateur doit donc pouvoir afficher un maximum d'informations dans un espace restreint, ce qui va motiver l'apparition de nouvelles formes lexicales :

Some are still in their infancy, with their situational status totally bound up with emerging technology, and therefore subject to rapid change: an example is the linking of the Internet to mobile phone technology, where the small screen size immediately motivated a fresh range of linguistic expression. (Crystal, 2001 : 10)

Certains médias sociaux, à l'image de la plateforme de microblogging *Twitter*, imposent par ailleurs un nombre limite de caractères par publication<sup>3</sup>, renforçant encore davantage ce besoin de recourir à de nouvelles expressions plus concises. L'on pensera également à la rapidité des échanges qui ne concerne désormais plus seulement les services de messageries instantanées, mais l'ensemble des médias sociaux du fait de la connexion rapide et constante dont bénéficient beaucoup d'utilisateurs. Ces derniers sont en capacité de publier en quelques secondes du contenu parce qu'ils réagissent sur le vif à quelque chose ou encore parce qu'ils prennent part à une conversation en ligne publiquement disponible.

<sup>2</sup> *Infomercial* désigne une publicité qui promeut un produit ou un service dans un style se voulant informatif et objectif (*OED*, en ligne).

<sup>3</sup> Pour *Twitter*, cette limite était de 140 caractères avant de passer à 280 caractères en novembre 2017 pour les langues utilisant des caractères latins.

Dans ce contexte, nous sommes amenés à penser que la communication par les médias sociaux favorisera naturellement l'utilisation, voire la création, de néologismes par abrègement<sup>4</sup>, dans la mesure où ils répondent à ce besoin de concision et de gain de place, permettant de réduire le temps nécessaire à la bonne transmission d'un message. En explorant le lien entre néologie par abrègement dans les médias sociaux et économie linguistique nous allons tenter vérifier si cette économie est réelle, du moins dans le contexte des médias sociaux.

### 3. Collecte de données

Cette étude s'appuie sur un corpus de 643 items collectés dans les médias sociaux. Pour cette collecte de néologismes, trois méthodes ont été utilisées.

La première consiste en une veille néologique sur des sources primaires : les néologismes sont collectés directement depuis différents médias sociaux. Cette méthode permet de collecter manuellement les néologismes directement dans leur contexte d'utilisation, mais elle présente l'inconvénient d'être chronophage et peu productive.

La seconde méthode de collecte consiste en un recours à des sources secondaires, à savoir des listes déjà établies de néologismes. Certains sites, à l'image d'*Urban Dictionary* ou encore *Wordspy*, proposent en effet ce genre de contenu. Cette méthode est moins chronophage et plus productive que la première, toutefois, il convient de s'assurer systématiquement de l'emploi réel de chaque néologisme dans les médias sociaux.

La troisième méthode consiste en une collecte automatique sur le réseau social *Twitter* au moyen d'un script en langage Python s'appuyant sur le module *Tweepy*<sup>5</sup>. Cette méthode offre d'une part la possibilité de collecter de nouveaux items en recherchant des tweets contenant des séquences du type « new portmanteau word », « coined a new word » ou encore « figure out an acronym » et permet d'autre part d'opérer une recherche de tweets contenant un ou des items particuliers (notamment pour vérifier le contexte d'utilisation d'un item donné ou bien pour comparer la fréquence d'emploi de deux variantes orthographiques). Les données collectées grâce à ce script sont ensuite compilées sous forme de tableur pour être facilement exploitées.

Dans la mesure où cette étude s'inscrit dans un projet plus large qui utilise les néologismes par abrègement comme support à un travail sur des questions dans les domaines de la morphologie et de phonologie de l'anglais, aucun seuil en termes de fréquence d'emploi du néologisme n'a été retenu car c'est bien le processus morphologique qui nous intéresse alors. Tout néologisme répondant à l'un des trois procédés de formation étudiés a donc été intégré au corpus.

## 4. Observations

### 4.1. Productivité

Une première observation peut être faite quant à la productivité (au sens de fréquence d'emploi) de chaque type de néologisme : sur 643 items, le corpus se constitue de 437 amalgames, 155 troncations et 51 acronymes. Au regard de ces données, on constate que l'amalgamation est un procédé apparemment beaucoup plus productif que la troncation et l'acronymie. Notons également que les items ont été collectés à partir de contenu de domaines très divers (politique, sport, mode, jeux vidéo, etc.). Notre corpus ne nous a pas permis

---

<sup>4</sup> Nous attirons l'attention sur le fait qu'il ne s'agit ici nullement de dire que l'économie linguistique est la seule motivation à l'emploi et à la création de ce type de néologisme.

<sup>5</sup> Ce module permet d'exploiter l'API (*Application Programming Interface* ou interface de programmation applicative) développée par Twitter qui rend possible la collecte de toutes sortes de données publiées par les utilisateurs du réseau social, notamment à des fins marketing et commerciales.

d'observer qu'un des trois procédés serait particulièrement productif dans du contenu relevant d'un domaine précis. Pour tenter d'expliquer cette différence de productivité à l'échelle de tout le corpus, il convient d'abord de noter que le contenu des médias sociaux présente à la fois des caractéristiques de l'écrit et de l'oral, ainsi que le note Crystal (2001) : « What makes Netspeak so interesting, as a form of communication, is the way it relies on characteristics belonging to both sides of the speech/writing divide » (: 28). Cela rend la frontière entre oral et écrit floue. Néanmoins, hormis sur les plateformes de partage de contenu vidéo, la communication passe avant tout par l'écrit.

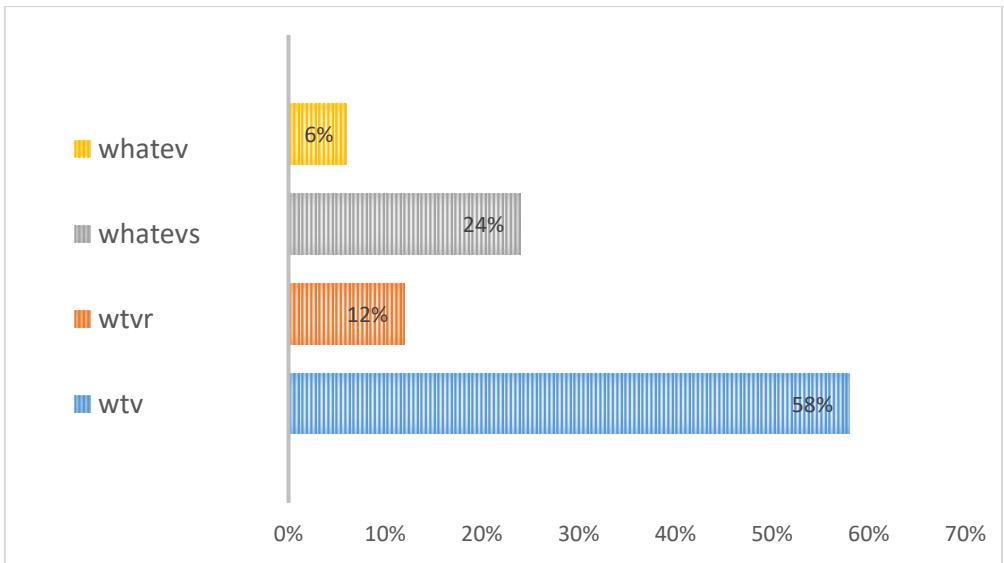
En ce qui concerne les acronymes, nous pouvons supposer que leur faible productivité est due au fait qu'ils sont en concurrence avec les sigles. En termes d'économie linguistique, si à l'oral l'intérêt de l'acronymie par rapport à la siglaison s'envisage volontiers (il est plus aisé et plus rapide de prononcer un mot que d'épeler une série d'initiales), rien ne contraint la préférence d'un acronyme par rapport à un sigle à l'écrit. Sur le plan cognitif, s'efforcer de créer une séquence d'initiales qui puisse être oralisée en jouant, soit sur la séquence de mots, soit sur les initiales retenues<sup>6</sup>, sera même plus fastidieux et engendrera un coût plus important que d'assembler les premières lettres de chaque mot d'une séquence quelconque de mots.

Un constat similaire peut être fait pour les troncations. Certes, la création et/ou l'emploi de troncations, phénomène avant tout oral, dans un média dont la communication se fait essentiellement par l'écrit peut sembler ne pas aller de soi, mais comme nous l'avons mentionné un peu plus tôt, il n'existe pas de frontière claire entre oral et écrit dans les médias sociaux : bien souvent, on écrit comme on parle. Toutefois, à l'image de ce que nous avons noté pour les acronymes et les sigles, l'utilisateur peut de nouveau choisir entre avoir recours à une troncation (qui, comme nous l'avons mentionné lorsque nous avons défini ce procédé, doit être prononçable pour être considéré comme une troncation puisqu'il s'agit d'abord d'un phénomène oral) ou bien à une abréviation, phénomène qui n'est possible qu'à l'écrit puisqu'il suffit de retenir quelques caractères nécessaires à la reconnaissance du mot source sans se soucier de rendre l'abréviation prononçable (ce qui permet, par exemple, de réduire un mot à ses seules consonnes). À la lecture, l'abréviation sera lue sous sa forme pleine. À nouveau, la création d'une abréviation demande un effort cognitif moindre que la création d'une troncation. Pour illustrer ce phénomène de concurrence, nous avons, grâce au module *Tweepy* mentionné en amont, entrepris une collecte automatique de tweets contenant une des quatre formes abrégées suivantes du mot *whatever* : deux abréviations, *wtvr* et *wtv*, et deux troncations *whatev* et *whatevs*<sup>7</sup>. Tous les tweets contenant une de ces variantes ont été collectés simultanément et en temps réel sur une période de trois jours, ce qui représente un total de 2106 tweets<sup>8</sup>. Sur les 2106 tweets collectés au cours de cette période, nous constatons les résultats suivants :

<sup>6</sup> À l'image de l'acronyme *SWEDOW* : *Stuff WE DON't Want*, séquence dont a parfois été retenue la deuxième lettre d'un élément à des fins d'oralisation.

<sup>7</sup> Notons la présence d'un *-s* hypocoristique à la fin de la deuxième forme tronquée. Ce phénomène est assez courant lors du processus de troncation, à l'image de cet autre exemple : *totally* → *totes*.

<sup>8</sup> Un grand merci à Michael Nauge, ingénieur d'étude au laboratoire FoReLLIS (EA 3816), pour son aide précieuse dans la mise en place de ce protocole.



**Figure 1 – Fréquences d'emploi de formes raccourcies du mot *whatever* sur 2106 tweets.**

On observe que l'abréviation *wtv* est largement favorisée par rapport aux trois autres puisqu'elle apparaît dans 58% des tweets collectés. La forme tronquée *whatevs* est la plus courante des deux troncations, toutefois, on ne la trouve que dans un peu moins d'un quart des tweets collectés. Il convient de préciser que ces résultats n'ont pour objectif que d'illustrer la possibilité pour un utilisateur de choisir entre troncation et abréviation. Il n'est nullement question de tenter d'affirmer de manière absolue que l'un ou l'autre de ces deux procédés est plus courant. On observe tout de même que, dans le cas de *whatever*, c'est la forme la plus brève (trois consonnes seulement) qui est privilégiée par les utilisateurs.

Les amalgames constituent quant à eux le procédé le plus productif de notre corpus. Pour expliquer cette productivité, nous envisageons deux hypothèses, complémentaires l'une de l'autre. La première est que l'amalgamation permet de répondre à un besoin de mots nouveaux (pour combler un vide lexical, par exemple) tout en réduisant la place prise par ces mots. Pour illustrer cela, nous pouvons citer l'exemple de « *she is smexting* », bien plus court que « *she is smoking while texting* », tout comme *hackathon* est plus court que *hacking marathon*. Dans ces deux exemples, l'amalgame permet en même temps de conceptualiser deux nouvelles pratiques. La seconde hypothèse pour expliquer la productivité des amalgames découle du caractère ludique de ce type de formation lexicale. Ce caractère permet d'attirer l'attention du public sur une publication, ce qui explique que le recours à l'amalgame soit fréquent dans la presse ou dans la publicité, ainsi que l'explique Lehrer (2003) :

« The perlocutionary intentions include using a word to catch the hearer's attention. This is especially important, since many neologisms, including blends, occur in print – ads, newspaper and magazine articles, etc. – where there is a great deal of competition for the reader's attention. Secondly, neologisms, especially in ads, are intended to be memorable. The speaker wants the hearer to remember the name of the product, process, or business establishment. »  
(Lehrer, 2003 : 370)

Nous pouvons émettre l'hypothèse que, de la même manière, dans le contexte des médias sociaux, l'emploi d'amalgames permet d'attirer l'attention sur un message, offrant à ce dernier une meilleure visibilité au sein de la quantité massive de contenu publié sur Internet.

## 4.2. Une économie linguistique toute relative

Si l'économie de place permise par les néologismes par abrègement est réelle, le lien entre ce type de formation et économie linguistique au sens large doit être nuancé. En ce qui concerne en premier lieu la question de l'économie de place, nous avons constaté lors de la collecte de nos données que le recours à ce type de néologismes était fréquent dans des messages pourtant déjà très courts. Sur Twitter cela s'illustre par des messages bien en dessous du nombre autorisé de caractères. Eisenstein (2013b) puis Reddy *et al.* (2014) avaient déjà fait le constat que sur Twitter les formes abrégées étaient principalement employées dans des tweets d'une longueur bien inférieure à 140 caractères, limite en vigueur jusqu'en novembre 2017, permettant à ces deux études de conclure que le recours à ce type de mot n'était à l'évidence pas contraint par la limite imposée. Selon nous, il serait pourtant un peu hâtif de conclure que ce type de limite ne joue aucun rôle dans l'utilisation d'abrègements. En effet, on pourrait aussi envisager que l'utilisateur, au fil de son utilisation de *Twitter*, ait intériorisé cette limite et ait pris l'habitude d'employer un vocabulaire plus concis pour s'assurer de ne pas rédiger un message trop long. De plus, quand bien même ils ne seraient pas nécessairement motivés par une économie de place ou de caractères, ils limitent le coût physique car, étant plus courts, ils nécessitent moins d'effort à la saisie que pour le mot ou la séquence source. Notons également que, parce qu'ils sont plus courts, le temps pour écrire un abrègement est réduit, ce qui peut permettre une communication plus fluide entre utilisateurs.

En revanche, si selon Tournier (1991), que nous avons cité plus haut, le fait de raccourcir les mots découlait d'une volonté de limiter l'effort « à ce qui est nécessaire et suffisant à la transmission d'un message » (1991 : 63), cette idée doit être relativisée, du moins dans le contexte des médias sociaux. Nous avons mentionné en amont que, comparé à l'abréviation et à la siglaison, le recours à la troncation et à l'acronymie demandait un effort cognitif supérieur à l'utilisateur. Du côté du receveur, il n'est pas toujours aisé de comprendre le sens d'un néologisme par abrègement puisqu'un effort supplémentaire pourra être nécessaire à son interprétation. Lehrer (2003), qui a conduit des expériences pour comprendre comment les lecteurs et auditeurs de l'anglais traitaient les néologismes, fait un constat semblable, en particulier pour les amalgames :

However, what is unusual about most new blends and other trendy neologisms is that they don't increase efficiency. In fact, they create more effort to interpret – at least at first, until readers and hearers have figured out what the source words are and what they mean. *Cocacolonization* (< *Coca cola* + *colonization*) is easy to get, but *squangle* (< *square* + *angle*) and *narcoma* (< *narcotic* + *coma*) are harder to process than the whole compounds. (Lehrer, 2003 : 1)

Ainsi, ils ne rendent pas nécessairement la transmission d'un message plus efficace et on trouve parfois des publications dans lequel l'utilisateur, après avoir employé un néologisme, en précise immédiatement la signification, à l'image de ce tweet publié le 18 août 2018 à 22h14 : « Lucky? Test that theory. I'm sure the guvment (government) will fund that study<sup>9</sup>. ». L'utilisateur entreprend d'éliminer toute ambiguïté qui nuirait à la bonne compréhension du message. On ne peut donc pas dire, dans ce cas précis, que l'emploi de la troncation est motivé par l'économie linguistique.

Pour nuancer encore le lien entre économie linguistique et ce type de néologisme, il est important de prendre en compte le fait que ces derniers dénotent souvent l'appartenance à un groupe social, ce que Plag (2003) notait déjà pour les troncations (en dehors du contexte des médias sociaux) :

<sup>9</sup> Lien vers la publication : <https://twitter.com/statuses/1030910846833238018> (vérifié le 15/05/2019).



Clippings appear as a rather mixed bag of forms abbreviated from larger words, which, however, share a common function, namely to express familiarity with the denotation of the derivative [...]. Thus, lab is used by people who work in laboratories, demo is part of the vocabulary of people who attend demonstrations, and so on. (Plag, 2003 : 121)

Ainsi, dans les médias sociaux, un joueur (de jeux vidéo) désigne un débutant par le mot *noob* (troncation de *newbie*), *ROFL* (prononcé /rɒfl/) est un acronyme pour *Rolling On the Floor Laughing* employé principalement des utilisateurs plutôt jeunes pour exprimer le fait que quelque chose les fait rire, et, dans les milieux féministes, *mansplain* (amalgame de *man* et *explain*) est un verbe utilisé pour désigner l'acte d'un homme qui explique à une femme ce qu'elle sait déjà de manière condescendante, partant du principe qu'il sait mieux qu'elle. Voici quelques exemples de ces mots en contexte :

- « I feel like a noob playing Xbox. It's been a year »<sup>10</sup>
- « Such a sick show with DZ Deathrays! Rofl I can't believe you remember the few haha! If you're looking for a place to play in Indy hit up MOKB and the @thehifiindy. »<sup>11</sup>
- « A man wants to help get women on boards. @MargaretSheil says to men, "don't mansplain!". Make sure women on boards are listened to and get airtime in meetings #getonboardQLD »<sup>12</sup>

Pour le non-initié, le sens de ces néologismes peut demeurer obscur. Un effort cognitif supplémentaire lui est nécessaire pour tenter de remonter aux mots sources et ainsi saisir le sens de ces néologismes qui font ainsi partie d'un véritable jargon propre à un groupe social. En conséquence du développement de ce type de jargon, des articles adressés à un public plus large désireux de décrypter les jargons des internautes sont régulièrement publiés. À titre d'exemples, nous mentionnerons les trois articles suivants :

- 28 Internet acronyms every parent should know (CNN, 09/01/2015)
- 30 Trendy Internet Slang Words and Acronyms You Need to Know to Fit In (MakeUseOf, 15/07/2015)
- Do you understand the teen slang on Ask.fm? (BBC News, 19/01/2016)

Si les premier et troisième articles sont clairement adressés à des adultes souhaitant comprendre le jargon employé par les plus jeunes sur Internet ; le second article, lui, entreprend d'enseigner quelques néologismes nécessaires pour bien être intégré à la communauté des internautes, ou du moins donner l'impression de faire partie de ce groupe.

Ce sentiment d'exclusion d'un groupe social parce qu'on n'en maîtrise pas le jargon est directement lié à l'effort cognitif supplémentaire que doit fournir l'individu devant un mot inconnu. Dans le cas qui nous intéresse, il s'agit de l'effort que doit fournir un individu pour retrouver le(s) mot(s) source(s) ou une séquence source. À l'inverse, pour celui qui en maîtrise le sens, le néologisme par abrègement permet effectivement une économie de l'effort cognitif puisqu'il s'agit d'une forme plus concise que son mot source, plus facile à mémoriser. Progressivement, une forme abrégée pourra même tendre à remplacer son mot source dans le lexique de l'utilisateur (voire dans le lexique commun), ou voir son sens se spécialiser par rapport au sens du mot source. On peut mentionner le cas de *fragmentation device* qui désigne à l'origine une grenade à fragmentation et dont la forme tronquée *frag* a donné dans le jargon militaire américain un verbe désignant le fait de lancer ce type de grenade sur un officier

<sup>10</sup> Lien vers la publication : <https://twitter.com/statuses/1034797589366009856> (vérifié le 15/05/2019).

<sup>11</sup> Lien vers la publication : <https://twitter.com/statuses/1034796262623731712> (vérifié le 15/05/2019).

<sup>12</sup> Lien vers la publication : <https://twitter.com/statuses/1034645260809396224> (vérifié le 15/05/2019).

supérieur manifestant trop d'ardeur au combat<sup>13</sup>, avant de donner, dans le jargon des joueurs de jeux vidéo de tir à la troisième personne, un nom qui désigne le fait pour un joueur de tuer son adversaire dans le jeu<sup>14</sup>.

## 5. Conclusion

Comme nous avons pu le voir, le lien entre économie linguistique, au sens entendu par Tournier (1991), et néologie par abrègement ne va pas nécessairement de lui-même. Si ces formes permettent une certaine économie du coût physique nécessaire à l'écriture, l'articulation ou, dans le cas de la communication médiée par ordinateur, la saisie d'un message, il n'en est pas toujours de même sur le plan cognitif. Le locuteur doit faire l'effort de trouver une forme qui répondra aux contraintes phonotactiques de l'anglais, contrairement à d'autres procédés propres à l'écrit tels que l'abréviation ou la siglaison qui permettent une économie de caractères, tout en n'étant pas soumis aux mêmes contraintes. Du côté du receveur, analyser et comprendre une forme abrégée lui demandera, s'il ne connaît pas cette forme, un effort mental supplémentaire, effet qui peut être recherché par l'utilisateur employant un jargon pour inclure et exclure des individus de sa communauté.

Nous terminerons en rappelant que la communication et la publication de contenu via Internet se font principalement en anglais. En conséquence, de nombreux utilisateurs parlant d'autres langues empruntent des formes anglophones (dont des néologismes par abrègement) pour les intégrer à leur langue, à l'image du terme *selfie* qui a intégré deux dictionnaires français en 2016. Le recours à la traduction est assez faible, et cela est d'autant plus vrai pour les néologismes par abrègement, car il y a alors la contrainte supplémentaire de respecter le procédé de formation lexicale du terme d'origine. Il faut en plus que ce processus de formation lexicale existe dans le système linguistique de la langue cible. Nous pensons qu'il sera intéressant pour d'éventuelles recherches futures de se pencher sur l'effort cognitif fourni par un sujet lors de la traduction de formations de type mais aussi lors de la réception (à l'écrit et à l'oral) de formes empruntées et leurs équivalents traduits (tels que *mansplaining* et sa traduction française *mecsplication*), afin de comprendre cette tendance au recours à l'emprunt plutôt qu'à la traduction.

## Références bibliographiques

- CRYSTAL, David (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- EISENSTEIN, Jacob (2013a). Phonological factors in social media writing, in *Proceedings of the workshop on Language Analysis in Social Media*: 11-19.
- EISENSTEIN, Jacob (2013b). What to do about bad language on the internet, in *Proceedings of HLT-NAACL*: 359-369.
- FISCHER, Roswitha (1998). *Lexical Change in Present Day English*. Tübingen: Gunter Narr.
- GARDIN, Bernard, LEFEBVRE, G., MARCELLESI, Christiane, MORTUREUX, Marie-Françoise (1974). À propos du « sentiment néologique », *Langages*, 36, 45-52.
- KAPLAN, Andrea & HAENLEIN, Michael (2010). Users of the world, unite! The challenges and opportunities of social media, in *Business Horizons*, 53(1): 59-68.
- LEHRER, Adrienne (2003). Understanding trendy neologisms, *Rivista di Linguistica*, 15(2): 369-382.
- PLAG, Ingo (2003). *Word-Formation in English*. Cambridge: Cambridge University Press.
- PRUVOST, Jean & SABLAYROLLES, Jean-François (2016). *Les néologismes*. Paris : Presses Universitaires de France.

<sup>13</sup> « To throw a fragmentation grenade at one's superior officer, esp. one who is considered over-zealous in his desire for combat. » (*OED*, en ligne).

<sup>14</sup> « Frag is a commonly used term in first-person shooter game. When you kill an opponent in the game you call it a "frag" rather than kill. » (*Webopedia* : <https://www.webopedia.com/TERM/F/frag.html>).

REDDY, Sravana, STANFORD, James & ZHONG, Joy (2014), A Twitter-Based Study of Newly Formed Clippings in American English, *Presentation at the Annual Meeting of the American Dialect Society*.

TOURNIER, Jean (1991). *Structures lexicales de l'anglais*. Paris : Nathan.

### ***Dictionnaires en ligne consultés***

*Merriam Webster*: <https://www.merriam-webster.com/>, (2019). Consulté en mai 2019

*Oxford English Dictionary*: <https://www.oed.com/>, (2019). Consulté en mai 2019

*Urban Dictionary*: <https://www.urbandictionary.com/>, (2019). Consulté en mai 2019

### ***Références aux sites Internet***

*BBC News*: <https://www.bbc.com/news>, (2019). Consulté en mai 2019

*CNN* : <https://edition.cnn.com/>, (2019). Consulté en mai 2019

*MakeUseOf*: <https://www.makeuseof.com/>, (2019). Consulté en mai 2019

*Twitter*: <https://twitter.com/>, (2019). Consulté en mai 2019

*Webopedia*: <https://www.webopedia.com/>, (2019). Consulté en mai 2019

## Avec le soutien de

